On the absence of the RIP in real-world applications of compressed sensing and the RIP in levels

Alexander Bastounis
CCA, Centre for Mathematical Sciences
University of Cambridge
Wilberforce Rd, Cambridge CB3 0WA
United Kingdom

Anders C. Hansen
DAMTP, Centre for Mathematical Sciences
University of Cambridge
Wilberforce Rd, Cambridge CB3 0WA
United Kingdom

Abstract

The purpose of this paper is twofold. The first is to point out that the Restricted Isometry Property does not hold in many applications where compressed sensing is successfully used. This includes some of the flagships of compressed sensing like Magnetic Resonance Imaging. We demonstrate that for natural compressed sensing matrices involving a level based reconstruction basis, e.g. wavelets, the number of measurements required to recover all s-sparse signals for reasonable s is excessive. In particular, uniform recovery of all s-sparse signals is quite unrealistic. This realisation shows that the Restricted Isometry Property is insufficient for explaining the success of compressed sensing in various practical applications. The second purpose of the paper is to introduce a new realistic framework based on a new RIP-like definition that fits the actual applications where compressed sensing is used. We show that the shortcomings used to show that uniform recovery is unreasonable no longer apply if we instead ask for structured recovery that is uniform only within each of the levels. To analyse this phenomenon, a new tool, termed the 'Restricted Isometry Property in Levels' is described and analysed. We show that with certain conditions on the Restricted Isometry Property in Levels, a form of uniform recovery within each level is possible. Finally, we conclude the paper by providing examples that demonstrate the optimality of the results obtained.

1 Introduction

Compressed Sensing (CS), introduced by Candès, Romberg & Tao [9] and Donoho [13], has been one of the major new developments in applied mathematics in the last decade [7, 12, 14–16, 18]. By introducing a non-linear reconstruction method via convex optimisation, one can circumvent traditional barriers for reconstructing vectors that are sparse or compressible, meaning that they have few non-zero coefficients or can be approximated well by vectors with few non-zero coefficients.

A substantial part of the CS literature has been devoted to the Restricted Isometry Property (RIP) (see Definition 1.2). Matrices which satisfy the RIP of order s (see Remark 1.5) can be used to perfectly recover all s-sparse vectors (i.e. vectors with at most s non-zero coefficients) via ℓ^1 minimization. Remarkably, this is possible even if the matrix in question is singular.

Given the substantial interest in the RIP over the last years it is natural to ask whether this intriguing mathematical concept is actually observed in many of the applications where CS is applied. It is well known that verifying that the RIP holds for a general matrix is an NP hard problem [21], however, there is a simple test that can be used to show that certain matrices do not satisfy the RIP. This is called the *flip test*. As this test reveals, there is an overwhelming amount of practical applications where one will not observe the RIP of order s for any reasonable sizes of the sparsity s. In particular, this list of applications contains some of the CS flagships such as Magnetic Resonance Imaging (MRI), other areas of medical imaging such as Computerised Tomography (CT) and all the tomography cousins such as thermoacoustic, photoacoustic or electrical impedance tomography, electron microscopy, seismic tomography, as well as other fields such as fluorescence microscopy, Hadamard spectroscopy and radio interferometry. Typically, we shall see that practical applications that exploit sparsity in a level based reconstruction basis will not exhibit the RIP of order s for reasonable s.

We will thoroughly document the lack of RIP of order s in this paper, and explain why it does not hold for reasonable s. It is then natural to ask whether there might be an alternative to the RIP that may be more suitable for the actual real world CS applications. With this in mind, we shall introduce the *RIP in levels* which generalises the classical RIP and is much better suited for the actual applications where CS is used.

1.1 Compressed sensing

We shall begin by discussing the general ideas of compressed sensing. Typically we are given a scanning device, represented by an invertible matrix $M \in \mathbb{C}^{n\times n}$, and we seek to recover information $x \in \mathbb{C}^n$ from observed measurements $y := Mx$. In general, we require knowledge of every element of y to be able to accurately recover x without additional structure. Indeed, let $\Omega = \{\omega_1, \omega_2, \dots, \omega_{|\Omega|}\}$ with $1 \le \omega_1 < \omega_2 < \omega_3 < \cdots < \omega_{|\Omega|} \le n$ and define the projection map $P_\Omega : \mathbb{C}^n \to \mathbb{C}^{|\Omega|}$ so that $P_\Omega(x_1, x_2, \dots, x_n) := (x_{\omega_1}, x_{\omega_2}, \dots, x_{\omega_{|\Omega|}})$. If $|\Omega|$ is strictly less than n then for a given y there are at least two distinct vectors $x_1 \in \mathbb{C}^n$ and $x_2 \in \mathbb{C}^n$ with $P_\Omega y = P_\Omega M x_1 = P_\Omega M x_2$, so that knowledge of $P_\Omega y$ will not allow us to distinguish between multiple possibilities of x. Ideally though we would like to be able to take $|\Omega| \ll n$ to reduce either the computational or financial costs associated with using the scanning device M.

So far, we have not assumed any additional structure on x. However, let us consider the case where the vector x consists mostly of zeros. More precisely, we make the following definition:

Definition 1.1 (Sparsity)**.** *A vector x is said to be* s-sparse *for some natural number s if* $|\mathrm{supp}(x)| \le s$*, where* $\mathrm{supp}(x)$ *denotes the support of the vector x (the indices corresponding to the non-zero values).*

The key to CS is the fact that, under certain conditions, finding solutions to the ℓ^1 problem

$$\min \|\widehat{x}\|_1 \text{ such that } P_\Omega M x = P_\Omega M \widehat{x} \tag{1.1}$$

gives a good approximation to x. More specifically, in [10] Candès and Tao described the concept of *Restricted Isometry Property*. Typically, M is an isometry. With this in mind, it may seem plausible that $U := P_\Omega M$ is also close to an isometry when acting on sparse vectors. Specifically, we make the following definition:

Definition 1.2 (Restricted Isometry Property)**.** *A matrix $U \in \mathbb{C}^{m\times n}$ is said to satisfy the* Restricted Isometry Property *(RIP) with* RIP Constant δ_s *if, for all s sparse vectors $x \in \mathbb{C}^n$*

$$(1-\delta_s)\|x\|_2^2 \le \|Ux\|_2^2 \le (1+\delta_s)\|x\|_2^2. \tag{1.2}$$

A typical theorem in compressed sensing is similar to the following, proven in [3] and [18].

Theorem 1.3. *Let $|\Omega| = m$, so that $P_\Omega y \in \mathbb{C}^m$ and let $U \in \mathbb{C}^{m\times n}$. Furthermore, suppose that U has RIP constant for $2s$-sparse vectors given by δ_{2s}, where $\delta_{2s} < \frac{4}{\sqrt{41}}$. Then for any $x \in \mathbb{C}^n$ such that $\|Ux - y\|_2 \le \epsilon$ and any $\tilde{x}$ which solve the following ℓ^1 minimization problem:*

$$\min \|\widehat{x}\|_1 \textit{ subject to } \|U\widehat{x} - P_\Omega y\|_2 \le \epsilon \tag{1.3}$$

we have

$$\|x - \tilde{x}\|_1 \le C\sigma_s(x)_1 + D\epsilon\sqrt{s} \tag{1.4}$$

and

$$\|x - \tilde{x}\|_2 \le \frac{C}{\sqrt{s}}\sigma_s(x)_1 + D\epsilon \tag{1.5}$$

where

$$\sigma_s(x)_1 := \min \|x - \widehat{x}_2\|_1 \textit{ such that } \widehat{x}_2 \textit{ is } s \textit{ sparse}$$

and C, D depend only on δ_{2s}.

Theorem 1.3 roughly says that if the matrix U has a sufficiently small RIP constant for $2s$-sparse vectors then

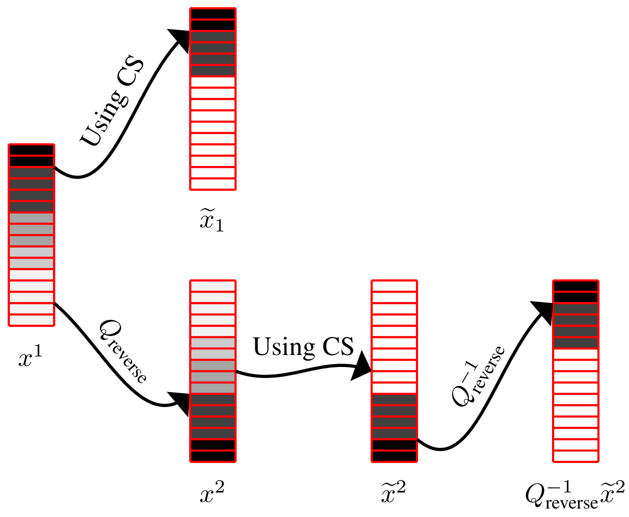

Figure 1: A graphical demonstration of the flip test for matrices which exhibit the RIP. Darker colours denote larger values.

1. If x is s-sparse then we can recover it from knowledge of Ux by solving the (convex) ℓ^1 minimization problem.

2. If we only know a noisy version of Ux then we can still recover an approximation to x which will be accurate up to the size of the noise.

3. If x is very close (in ℓ^1 norm) to an s-sparse vector then we can recover an approximation to x which will be accurate up to the distance between x and its closest s-sparse vector.

The ability to recover all s-sparse vectors x using the values of Ux is commonly known as *uniform recovery* . We shall see in the remaining sections that expecting uniform recovery is unrealistic in a variety of situations.

Remark 1.4. Variants on Theorem 1.3 include changing the hypothesis to $\delta_{2s} < C$ for a different constant C (e.g. [8] requires $\delta_{2s} < \sqrt{2} - 1$ or [17] uses different methods to show that, for large s, $\delta_{2s} < 0.475$ is sufficient) or changing δ_{2s} to δ_{3s} (e.g. [5] requires $\delta_{3s} < 4 - 2\sqrt{3}$) or δ_s (e.g. [6] requires $\delta_s < \frac{1}{3}$). We will see that any of these modifications will suffer from the same shortcomings in the next section.

Remark 1.5. If the RIP constant for $U \in \mathbb{C}^{m\times n}$ of order as for some $a \in \mathbb{N}$ is sufficiently small so that the conclusion of a Theorem similar to Theorem 1.3 holds then we say that U *satisfies or exhibits the RIP of order* s.

2 The absence of the RIP and the flip test

2.1 The flip test

Although Theorem 1.3 and similar theorems based on the RIP seem convenient, computing the RIP constant of an arbitrary matrix is an NP hard problem [21]. Certain special cases have been shown to have a small enough RIP constant with high probability (e.g. [11] for Gaussian and Bernoulli matrices) for the conclusion of Theorem 1.3 to hold but in other cases the size of the RIP constant is not known. In [2] the following test (the so called 'flip test') was proposed. The test can be used to verify the RIP of order s for reasonable s and uniform recovery of s-sparse vectors and works as follows:

1. Suppose that a matrix $U \in \mathbb{C}^{m\times n}$ has a sufficiently small RIP constant so that the error estimate (1.4) holds. Take a vector x^1 and compute $y^1 := Ux^1$. Set $\widetilde{x}^1$ to be the solution to (1.3) where we seek to reconstruct x^1 from y^1.

2. For an operator Q that permutes entries of vectors in $\mathbb{C}^n$, set $x^2 := Qx^1$, and compute $y^2 := Ux^2$. Again, set $\widetilde{x}^2$ to be the solution to (1.3) where we seek to reconstruct x^2 from y^2.

		Matrix method		Observes RIP
		$\mathrm{DFT} \cdot \mathrm{DWT}^{-1}$	$\mathrm{HAD} \cdot \mathrm{DWT}^{-1}$	
Problem	MRI	✓	✗	✗
	Tomography	✓	✗	✗
	Spectroscopy	✓	✗	✗
	Electron microscopy	✓	✗	✗
	Radio interferometry	✓	✗	✗
	Fluorescence microscopy	✗	✓	✗
	Lensless camera	✗	✓	✗
	Single pixel camera	✗	✓	✗
	Hadamard spectroscopy	✗	✓	✗

Table 1: A table displaying various applications of compressive sensing. For each application, a suitable matrix is suggested along with information on whether or not that matrix has a sufficiently small RIP constant for Theorems similar to Theorem 1.3 to hold.

3. From (1.4), we have
$$\|x^1 - \widetilde{x}^1\|_1, \|x^2 - \widetilde{x}^2\|_1 \le C\sigma_s(x^2)_1 = C\sigma_s(x^1)_1$$
because σ_s is independent of ordering of the coefficients. Also, since permutations are isometries,
$$\|x^1 - Q^{-1}\widetilde{x}^2\|_1 \le C\sigma_s(x^2)_1 = C\sigma_s(x^1)_1.$$
4. If x^1 is well approximated by $\widetilde{x}^1$ and the RIP is the reason for the good error term, we expect $\|x^1 - \widetilde{x}^1\|_1 \approx C\sigma_s(x^1)_1$. Thus, x^1 should also be well approximated by $Q^{-1}\widetilde{x}^2$.
5. Hence, if
$$\|x^1 - \widetilde{x}^1\|_1 \text{ differs greatly from } \|x^1 - Q^{-1}\widetilde{x}^2\|_1,$$
we cannot have the RIP, since there is no uniform recovery.

The particular choice of Q that was given in [2] was the permutation Q_{reverse} that reverses order - namely, if $x \in \mathbb{C}^n$ then
$$Q_{\text{reverse}}(x_1, x_2, \dots, x_{n-1}, x_n) = (x_n, x_{n-1}, \dots, x_2, x_1).$$
A graphical demonstration of the flip test if the RIP holds for typical x^1 with Q_{reverse} is given in Figure 1. Here, larger values are represented by darker colours, so that the first few values of x^1 are the largest. This is what we would expect if x^1 is the wavelet coefficients of a typical image.

This test is simple, and performing it with the permutation Q_{reverse} on a variety of examples, even with the absence of noise (see Figure 2), gives us a collection of problems for which the RIP does not account for the excellent reconstruction observed by experimental methods. In any of these cases, the RIP cannot explain why we are seeing such good results. For several other examples with flip tests that verify the lack of RIP in many of the fields that are considered flagships of compressed sensing in application, see [1] and [20]. The conclusion of the flip test for practical compressed sensing is summarized in Table 1.

It is worth noting that even with 97% sampling as in the second row of Figure 2, the RIP constant δ_s (for reasonable s) is still too large for the conclusion of a Theorem like 1.3. Although this may seem surprising at first, it is actually very natural when one understands how poor the recovery of the finer wavelet levels is for subsampling of matrices like $\mathrm{DFT} \cdot \mathrm{DWT}_4^{-1}$.

2.2 Why don't we see the RIP?

To illustrate the reason for the lack of RIP in the previous examples, let us consider a matrix $M \in \mathbb{C}^{n \times n}$ such that
$$M = \begin{pmatrix} M_1 & 0 & \dots & 0 & 0 \\ 0 & M_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & M_{J-1} & 0 \\ 0 & 0 & \dots & 0 & M_J \end{pmatrix},$$

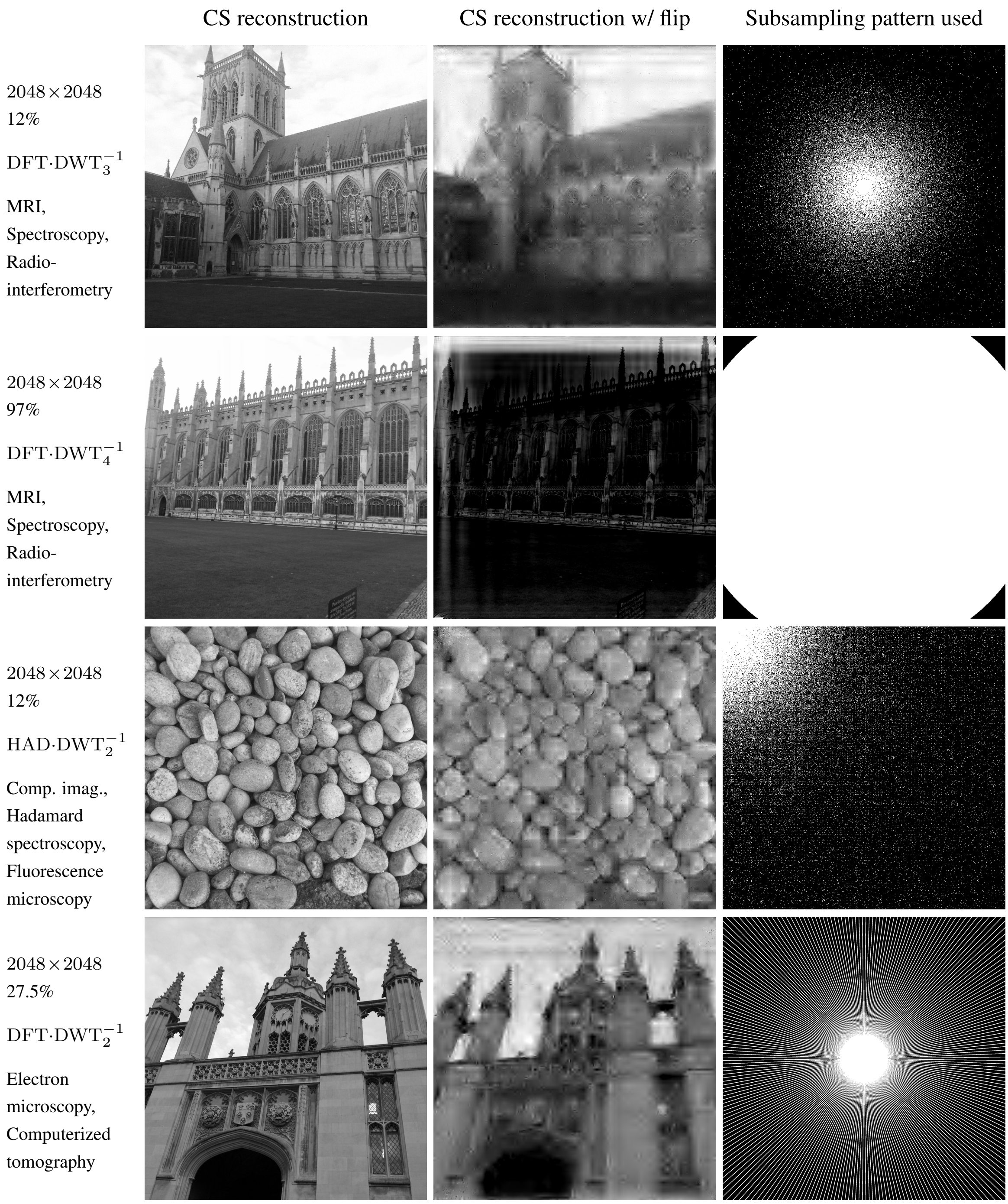

Figure 2: Standard reconstruction based on $\tilde{x}^1$ (left column) and the result from the flip test based on $Q_{\text{reverse}}\tilde{x}^2$ (middle column). The right column shows the subsampling pattern used (e.g. for the Fourier matrix the frequencies at which measurements were taken with the zero frequency at the centre of the image). The percentage shown is the fraction of Fourier or Hadamard coefficients that were sampled. DWT_N^{-1} denotes the Inverse Wavelet Transform corresponding to Daubechies wavelets with N vanishing moments.

Figure 3: The three images display the absolute values of various sensing matrices. A lighter colour represents larger absolute values. (Left) $\mathrm{DFT} \cdot \mathrm{DWT}_2^{-1}$, (middle) $\mathrm{DFT} \cdot \mathrm{DWT}_{10}^{-1}$ and (right) $\mathrm{HAD} \cdot \mathrm{DWT}_{\mathrm{Haar}}^{-1}$ where DFT is the Discrete Fourier Transform, HAD the Hadamard transform and $\mathrm{DWT}_{\mathrm{N}}^{-1}$ the Inverse Wavelet Transform corresponding to Daubechies wavelets with N vanishing moments.

where the matrices $M_i \in \mathbb{C}^{a_i} \times \mathbb{C}^{a_i}$ are isometries and a_i is the length of the i-th wavelet level. We shall consider the standard compressed sensing problem of recovering $x \in \mathbb{C}^n$, where $n = a_1 + a_2 + \cdots + a_J$, from knowledge of $P_\Omega M x$ for some sampling set Ω with $|\Omega| \ll n$. Given the block diagonality of M it is natural to consider a subsampling scheme that fits this structure. In particular, let

$$\Omega = \Omega_1 \cup \ldots \cup \Omega_J, \qquad \Omega_j \in \left\{ \sum_{k=1}^{j-1} a_k + 1, \ldots, \sum_{k=1}^{j} a_k \right\}, \quad 2 \leq j \leq J,$$

and $\Omega_1 \subset \{1, \ldots, a_1\}$. Also, let

$$P_\Omega = P_{\Omega_1} \oplus \ldots \oplus P_{\Omega_J},$$

where the direct sum refers to the obvious separation given by the structure of M. Note that the block diagonal structure of M is a simplified model of the real world, however, as Figure 3 demonstrates, it is a good approximation when considering the popular matrices $\mathrm{DFT} \cdot \mathrm{DWT}^{-1}$ and $\mathrm{HAD} \cdot \mathrm{DWT}^{-1}$. Here DFT denotes the discrete Fourier transform, HAD the Hadamard transform and DWT the discrete wavelet transform. In fact, $\mathrm{HAD} \cdot \mathrm{DWT}_{\mathrm{Haar}}^{-1}$ is actually completely block diagonal. The simplest example is when $J = 2$. The general idea follows from this simple model. Suppose that

$$x = (x^1, x^2), \quad x^1 = (x_1^1, \ldots, x_{a_1}^1), \quad x^2 = (x_1^2, \ldots, x_{a_2}^2), \quad a_1, a_2 \in \mathbb{N},$$

and

$$s_1 = |\mathrm{supp}(x^1)|, \quad s_2 = |\mathrm{supp}(x^2)|.$$

Given that it is natural to consider wavelets as the recovery basis, we will have that

$$s_2 \ll s_1,$$

(see Section 3.1 for a discussion of this phenomenon). Suppose now that $P_{\Omega_1} M_1$ and $P_{\Omega_2} M_2$ actually satisfy the RIP with reasonable constants $\delta_{s_1}^1 \approx \delta_{s_2}^2$ respectively. In this case we can easily recover x from $P_\Omega M x$ using ℓ^1 recovery. If the global RIP is the explanation for the success of compressed sensing, we would require $\delta_{s_1+s_2}$ to be of reasonable size for $P_\Omega M$, so that we can recover $\tilde{x} = Q_{\text{reverse}} x$ from $P_\Omega M \tilde{x}$. This leads to problems: since $s_2 \ll s_1$ we may have $\delta_{s_1}^1 \approx \delta_{s_2}^2 \ll \delta_{s_1}^2$.

As $\delta_{s_1+s_2} \geq \delta_{s_1}^2$, it is conceivable that the global RIP constant $\delta_{s_1+s_2}$ is huge despite the relatively small magnitude of both $\delta_{s_1}^1$ and $\delta_{s_2}^2$. The only conclusion is that despite our ability to recover using $P_\Omega M$, there is no reason to suspect that the global RIP constant for $P_\Omega M$ is sufficiently small for a result like Theorem 1.3 to apply.

Without equally good recovery on each wavelet level, there is no reason to suspect that the RIP constant corresponding to the entire matrix will be small. This is unrealistic - the levels corresponding to coarser details of an image (the more 'important' wavelet levels) typically exhibit better recovery than the finer detail wavelet levels. Even though the RIP is unable to explain the success of compressed sensing in Figure 2, it is clear that with these examples compressed sensing is indeed successful. We would like to emphasize the point that the RIP is merely a sufficient but not necessary condition for compressed sensing.

Remark 2.1. Although this example is simplified, both regarding the sampling procedure and the number of wavelet levels, similar arguments will explain the lack of RIP for different variable density sampling schemes. The key is that successful variable density sampling schemes will depend on the distribution of coefficients, just like this simplified example does.

3 A new theory for compressed sensing

The current mathematical theory for compressive sensing revolves around a few key ideas. These are the concepts of sparsity, incoherence, uniform subsampling and the RIP. In [1] and [20], it was shown that these concepts are absent for a large class of compressed sensing problems. To solve this problem, the new concepts of *asymptotic sparsity*, *asymptotic incoherence* and *multi-level sampling* were introduced. We now introduce the fourth new concept in the new theory of compressive sensing: the ***RIP in levels***. This generalisation of the RIP replaces the idea of *uniform recovery of s-sparse signals* with that of *uniform recovery of (s, M)-sparse signals*. We shall detail this new idea in this section.

3.1 A level based alternative: the RIP in levels

The examples given in Figure 2 all involve reconstructing in a basis that is divided into various levels. It is this level based structure that allows us to introduce a new concept which we term the 'RIP in levels'. We shall demonstrate such a concept based on our observations with wavelets, which we give a brief description of in the following section. Despite our focus on wavelets in the next few pages, it should be noted that our work applies equally to all level based reconstruction bases.

3.1.1 Wavelets

A multiresolution analysis (as defined in [19]) for $L^2(X)$ (where X is an interval or a square) is formed by constructing scaling spaces $(V_j)_{j=0}^{\infty}$ and wavelet spaces $(W_j)_{j=0}^{\infty}$ with $V_j, W_j \subset L^2(X)$ so that

1. For each $j \geq 0$, $V_j \subset V_{j+1}$.
2. If $f(\cdot) \in V_j$ then $f(2\cdot) \in V_{j+1}$, and vice-versa.
3. $\overline{\bigcup_{j=0}^{\infty} V_j} = L^2(X)$.
4. $\bigcap_{j=0}^{\infty} V_j = \{0\}$.
5. W_j is the orthogonal complement of V_j in V_{j+1}.

For natural images f, the largest coefficients in the wavelet expansion of f appear in the levels corresponding to smaller j. Closer examination of the relative sparsity in each level also reveals a pattern: let w be the collection of wavelet coefficients of f and for a given level k let S^k be the indices of all wavelet coefficients of f in the k-th level. Additionally, let $\mathcal{M}_n$ be the largest (in absolute value) n wavelet coefficients of f. Given $\epsilon \in [0, 1]$, we define the functions $s(\epsilon)$ and $s_k(\epsilon)$ (as in [1]) by

$$s(\epsilon) := \min \left\{ n : \left\| \sum_{i \in \mathcal{M}_n} w_i \right\| \geq \epsilon \left\| \sum_{k=0}^{\infty} \sum_{i \in S^k} w_i \right\| \right\}, \quad s_k(\epsilon) := |\mathcal{M}_{s(\epsilon)} \cap S^k|.$$

More succinctly, $s_k(\epsilon)$ represents the relative sparsity of the wavelet coefficients of f at the kth scale. If an image is very well represented by wavelets, we would like $s_k(\epsilon)$ to be small for as large ϵ as possible. However, one can make the following observation: if we define the numbers M_k so that $M_k - M_{k-1} = |S^k|$ (the reason for this choice of notation will become clear in Section 3.1.2) then the ratios $\frac{s_k(\epsilon)}{M_k - M_{k-1}}$ decay very rapidly for a fixed ϵ. Numerical examples showing this phenomenon with Haar Wavelets are displayed in Figure 4. The point is that the sparsity has a structure which the traditional RIP ignores.

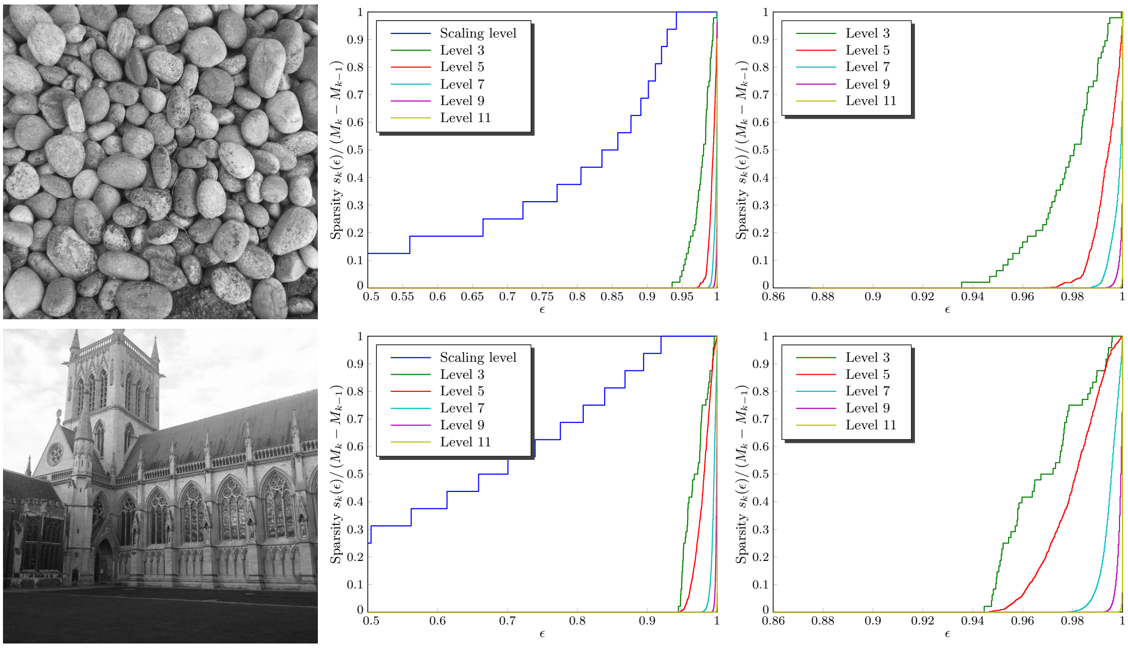

Figure 4: The relative sparsity of Haar wavelet coefficients of two image. The leftmost column displays the image in question. The middle and final columns display the values of $s_k(\epsilon)$ for $\epsilon \in [0.5, 1]$ and $\epsilon \in [0.86, 1]$ respectively, where k represents a wavelet level. Of particular importance is the rapid decay of $s_k(\epsilon)$ as k grows larger.

3.1.2 $(\mathbf{s}, \mathbf{M})$-sparsity and the flip test in levels

Theorem 1.3 and any similar theorems all suggest that we are able to recover all s-sparse vectors exactly, independent of which levels the s-sparse vectors are supported on. Instead of such a stringent requirement, we can take advantage of the structure of our problem. We have observed that, for wavelets, as $k \to \infty$ $s_k(\epsilon)/(M_k - M_{k-1})$ decays rapidly (see Figure 4). To further understand this phenomenon, in [1] the concept of $(\mathbf{s}, \mathbf{M})$-sparsity was introduced.

Definition 3.1 ($(\mathbf{s}, \mathbf{M})$-sparsity)**.** *Let*

$$\mathbf{M} = (M_0, M_1, \dots, M_l) \in \mathbb{N}^{l+1}$$

with $1 \leq M_1 < M_2 < \cdots < M_l$ and $M_0 = 0$, where the natural number l is called the number of levels. *Additionally, let*

$$\mathbf{s} = (s_1, s_2, \dots, s_l) \in \mathbb{N}^l,$$

with $s_i \leq M_i - M_{i-1}$. We call $(\mathbf{s}, \mathbf{M})$ a sparsity pattern. *A set Λ of integers is said to be $(\mathbf{s}, \mathbf{M})$-sparse if*

$$\Lambda \subset \{M_0 + 1, M_0 + 2, \dots, M_l\}$$

and for each $i \in \{1, 2, \dots, l\}$,

$$|\Lambda \cap \{M_{i-1} + 1, M_{i-1} + 2, \dots, M_i\}| \leq s_i.$$

A vector is said to be $(\mathbf{s}, \mathbf{M})$-sparse if its support is an $(\mathbf{s}, \mathbf{M})$-sparse set. The collection of $(\mathbf{s}, \mathbf{M})$-sparse vectors is denoted by $\Sigma_{\mathbf{s},\mathbf{M}}$.

Remark 3.2. If $(\mathbf{s}, \mathbf{M})$ is a sparsity pattern, we will sometimes refer to $(a\mathbf{s}, \mathbf{M})$-sparse sets for some natural number a even though as_i may be larger than $M_i - M_{i-1}$. To make sense of such a statement, we define (in this context)

$$a\mathbf{s} := (\min(as_1, M_1 - M_0), \min(as_2, M_2 - M_0), \dots, \min(as_l, M_l - M_{l-1}))\,.$$

We can also define $\sigma_{\mathbf{s},\mathbf{M}}(x)_1$ as a natural extension of $\sigma_s(x)_1$. Namely,

$$\sigma_{\mathbf{s},\mathbf{M}}(x)_1 := \min_{\widehat{x} \in \Sigma_{\mathbf{s},\mathbf{M}}} \|x - \widehat{x}\|_1.$$

Let us now look at a specific case where $(\mathbf{s}, \mathbf{M})$ represents wavelet levels (again, we emphasize that wavelets are simply one example of a level based system and that our work is more general). Roughly speaking, we can choose $\mathbf{s}$ and $\mathbf{M}$ such that x is $(\mathbf{s}, \mathbf{M})$-sparse if it has fewer non-zero coefficients in the finer wavelet levels. As in Theorem 1.3 (where we take $\epsilon = 0$ corresponding to an absence of noise), we shall examine solutions $\tilde{x}$ to the problem

$$\min \|\widehat{x}\|_1 \text{ such that } Ux = U\widehat{x}. \tag{3.1}$$

Instead of asking for

$$\|x - \tilde{x}\|_1 \leq C\sigma_s(x)_1$$

and

$$\|x - \tilde{x}\|_2 \leq C\frac{\sigma_s(x)_1}{\sqrt{s}},$$

we might instead look for a condition on U that allows us to conclude that

$$\|x - \tilde{x}\|_1 \leq C\sigma_{\mathbf{s},\mathbf{M}}(x)_1, \tag{3.2}$$

and

$$\|x - \tilde{x}\|_2 \leq \frac{C\sigma_{\mathbf{s},\mathbf{M}}(x)_1}{\sqrt{s_1 + s_2 + \cdots + s_l}}, \tag{3.3}$$

for some constant C, independent of x and $s, \eta_{\mathbf{s},\mathbf{M}}$. In Section 2.1, we saw that there was a simple test that was able to tell us if a matrix does not exhibit the RIP. However, the argument in Section 2.1 does not hold if we only insist on results of a form similar to equations (3.2) and (3.3). Instead, we can describe a 'flip test in levels' in the following way:

1. Suppose that a matrix $U \in \mathbb{C}^{m \times n}$ satisfies (3.2) and (3.3) when performing an ℓ^1 minimization of the same form as (3.1). Take a sample vector x^1 and compute $y^1 := Ux^1$. Set $\widetilde{x}^1$ to be the solution to (3.1) where we seek to reconstruct x^1 from y^1.

2. For an operator Q which permutes entries of vectors in $\mathbb{C}^n$ such that

$$Q(\Sigma_{\mathbf{s},\mathbf{M}}) = \Sigma_{\mathbf{s},\mathbf{M}} \tag{3.4}$$

   set $x^2 := Qx^1$, and compute $y^2 := Ux^2$. Again, set $\widetilde{x}^2$ to be the solution to (3.1) where we seek to reconstruct x^2 from y^2.

3. Equation (3.4) implies that $C\sigma_{\mathbf{s},\mathbf{M}}(x^1)_1 = C\sigma_{\mathbf{s},\mathbf{M}}(x^2)_1$, because

$$\begin{aligned}
\sigma_{\mathbf{s},\mathbf{M}}(x^2)_1 &= \min_{\widehat{x}^2 \in \Sigma_{\mathbf{s},\mathbf{M}}} \|x^2 - \widehat{x}^2\|_1 \\
&= \min_{\widehat{x}^2 \in Q(\Sigma_{\mathbf{s},\mathbf{M}})} \|x^2 - \widehat{x}^2\|_1 \\
&= \min_{\widehat{x}^1 \in \Sigma_{\mathbf{s},\mathbf{M}}} \|x^2 - Q\widehat{x}^1\|_1 \\
&= \min_{\widehat{x}^1 \in \Sigma_{\mathbf{s},\mathbf{M}}} \|Qx^1 - Q\widehat{x}^1\|_1 \\
&= \min_{\widehat{x}^1 \in \Sigma_{\mathbf{s},\mathbf{M}}} \|x^1 - \widehat{x}^1\|_1 = \sigma_{\mathbf{s},\mathbf{M}}(x^1)_1
\end{aligned}$$

   where we have used the fact that permutations are isometries in the final line.

4. From (3.2), we have

$$\|x^2 - \widetilde{x}^2\|_1 \leq C\sigma_{\mathbf{s},\mathbf{M}}(x^1)_1$$

   Again, since permutations are isometries,

$$\|x^1 - Q^{-1}\widetilde{x}^2\|_1 \leq C\sigma_{\mathbf{s},\mathbf{M}}(x^1)_1.$$

Table 2: Flip test in levels with randomly generated permutations

Image	Subsampling percentage	Max	Min	Standard deviation
College 1	12.48%	4.0560%	3.8955%	0.0642%
College 2	97.17%	0.4991%	0.4983%	0.0001%
Rocks	12.48%	5.6246%	5.5785%	0.0118%
College 3	27.51%	2.0074%	1.9860%	0.0032%

A table displaying relative error percentages $\|x-\widetilde{x}\|_2/\|x\|_2$ for various images x as in Figure 5 with recovered image $\widetilde{x}$. Each image was processed with a fixed subsampling pattern and 1000 randomly generated permutations as described in Section 3.1.2. The columns labelled 'Max', 'Min' and 'Standard deviation' list the maximum, minimum and standard deviation of the relative errors taken over all tested permutations.

5. If x^1 is well approximated by $\widetilde{x}^1$ and equations (3.2) and (3.3) are in some sense optimal, we expect $\|x^1 - \widetilde{x}^1\|_1 \approx C\sigma_{\mathbf{s},\mathbf{M}}(x^1)_1$. Thus, x^1 should also be well approximated by $Q^{-1}\widetilde{x}^2$.
6. If this fails, then equations (3.2) and (3.3) cannot be the reason for the success of compressed sensing with the matrix U.

Condition (3.4) now requires us to consider different permutations than a simple reverse permutation as in Section 2.1. A natural adaptation of Q_{reverse} to this new 'flip test in levels' is a permutation that just reverses coefficients within each wavelet level. Figure 5 displays what happens when we attempt to do the flip test with this permutation. In this case, we see that the performance of CS reconstruction under flipping and the performance of standard CS reconstruction are very similar. This suggests that uniform recovery within the class of $(\mathbf{s},\mathbf{M})$-sparse vectors (as in 3.2 and 3.3) is possible with a variety of practical compressive sensing matrices. Indeed, in Table 2 we also consider a collection of randomly generated Q for which equation (3.4) holds. We see that once again, performance with permutations within the levels is similar to standard CS performance.

3.1.3 The RIP in levels

Given the success of the 'flip test in levels', let us now try to find a sufficient condition on a matrix $U \in \mathbb{C}^{m\times n}$ that allows us to conclude (3.2) and (3.3). If the RIP implies (1.4) then the obvious idea is to extend the RIP to a so-called 'RIP in levels', defined as follows:

Definition 3.3 (RIP in levels)**.** *We say that U satisfies the* RIP in levels *(*RIP_L*) with* RIP_L *constant* $\delta_{\mathbf{s},\mathbf{M}} \geq 0$ *and sparsity pattern* $(\mathbf{s},\mathbf{M})$ *if for all x in $\Sigma_{\mathbf{s},\mathbf{M}}$ we have*

$$(1-\delta_{\mathbf{s},\mathbf{M}})\|x\|_2^2 \leq \|Ux\|_2^2 \leq (1+\delta_{\mathbf{s},\mathbf{M}})\|x\|_2^2.$$

We will see that the RIP in levels allows us to obtain error estimates on $\|x-\widetilde{x}\|_1$ and $\|x-\widetilde{x}\|_2$.

4 Main results

If a matrix $U \in \mathbb{C}^{m\times n}$ satisfies the RIP then we have control over the values of $\|Ue_i\|_2$ where $i \in \{1,2,\dots,n\}$ and e_i is the i-th standard basis element of $\mathbb{C}^n$. To ensure that the same thing happens with the RIP_L we make the following two definitions:

Definition 4.1 (Ratio constant)**.** *The* ratio constant *of a sparsity pattern* $(\mathbf{s},\mathbf{M})$*, which we denote by* $\eta_{\mathbf{s},\mathbf{M}}$*, is given by*

$$\eta_{\mathbf{s},\mathbf{M}} := \max_{i,j} \frac{s_i}{s_j}.$$

If there is a j for which $s_j = 0$ then we write $\eta_{\mathbf{s},\mathbf{M}} = \infty$.

Definition 4.2. *A sparsity pattern* $(\mathbf{s},\mathbf{M})$ *is said to* cover *a matrix* $U \in \mathbb{C}^{m\times n}$ *if*

CS reconstruction | CS reconstruction w/ flip in levels | Subsampling pattern used

College 1
2048×2048
12%
$\mathrm{DFT}\cdot\mathrm{DWT}_3^{-1}$
MRI, Spectroscopy, Radio-interferometry

College 2
2048×2048
97%
$\mathrm{DFT}\cdot\mathrm{DWT}_4^{-1}$
MRI, Spectroscopy, Radio-interferometry

Rocks
2048×2048
12%
$\mathrm{HAD}\cdot\mathrm{DWT}_2^{-1}$
Comp. imag., Hadamard spectroscopy, Fluorescence microscopy

College 3
2048×2048
27.5%
$\mathrm{DFT}\cdot\mathrm{DWT}_3^{-1}$
Electron microscopy, Computerized tomography

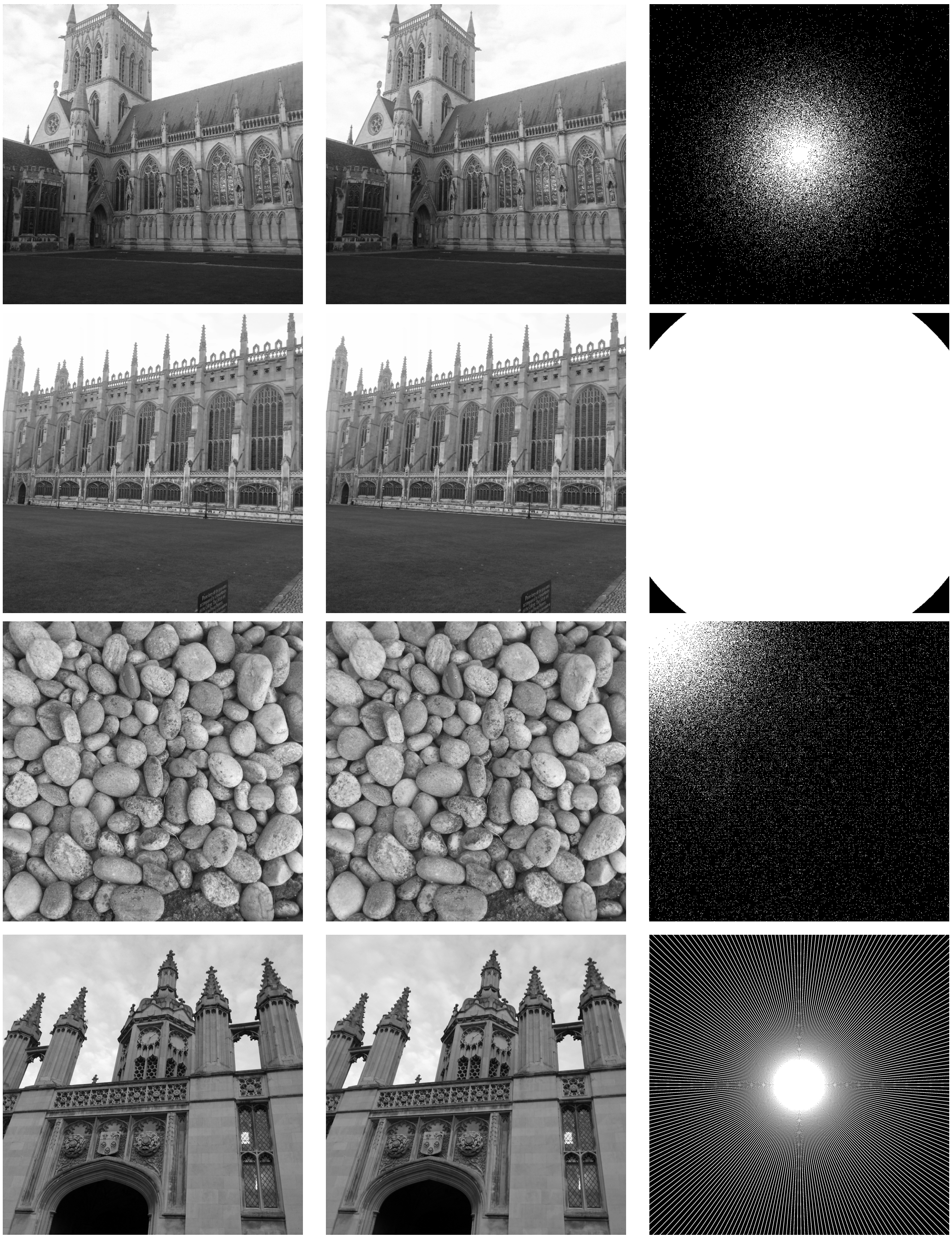

Figure 5: Standard reconstruction (left column) and the result of the flip test where the wavelet coefficients are flipped within each level (middle column). The right column shows the subsampling pattern used. The percentage shown is the fraction of Fourier or Hadamard coefficients that were sampled. DWT_N^{-1} denotes the Inverse Wavelet Transform corresponding to Daubechies wavelets with N vanishing moments.

1. $\eta_{\mathbf{s},\mathbf{M}} < \infty$

2. $M_l \geq n$ *where* l *is the number of levels for* $(\mathbf{s}, \mathbf{M})$.

If a sparsity pattern does not cover U because it fails to satisfy either 1 or 2 from the definition of a sparsity pattern covering a matrix U then we cannot guarantee recovery of $(\mathbf{s}, \mathbf{M})$-sparse vectors, even in the case that $\delta_{\mathbf{s},\mathbf{M}} = 0$. We shall justify the necessity of both conditions using two counterexamples. Firstly, we shall provide a matrix U, a sparsity pattern $(\mathbf{s}, \mathbf{M})$ and an $(\mathbf{s}, \mathbf{M})$ vector $x_1 \in \mathbb{C}^n$ such that $\eta_{\mathbf{s},\mathbf{M}} = \infty$, $\delta_{\mathbf{s},\mathbf{M}} = 0$ and x_1 is not recovered by standard ℓ^1 minimization. Indeed, consider the following

$$U = \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{s} = (1, 0), \quad \mathbf{M} = (0, 1), \quad x_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

By the definition of $\eta_{\mathbf{s},\mathbf{M}}$, we have $\eta_{\mathbf{s},\mathbf{M}} = \infty$ and it is obvious that $\delta_{\mathbf{s},\mathbf{M}} = 0$. Furthermore, even without noise, x_1 is not a minimizer of

$$\min \|\tilde{x}\|_1 \text{ such that } Ux_1 = U\tilde{x}.$$

because $Ux_1 = Ux_2$ with $\|x_2\|_1 = \frac{1}{2}$ where

$$x_2 := \begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix}.$$

It is therefore clear that Assumption 1 is necessary. We shall now provide a justification for Assumption 2 is also a requirement if we wish for the RIP_L to be a sufficient condition for recovery of $(\mathbf{s}, \mathbf{M})$-sparse vectors. This time, consider the following combination of U, $(\mathbf{s}, \mathbf{M})$ and x_1:

$$U = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{s} = (1, 1), \quad \mathbf{M} = (0, 1), \quad x_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

and again, even though $\delta_{\mathbf{s},\mathbf{M}} = 0$, recovery is not possible because $Ux_1 = Ux_2$ with $\|x_2\|_1 = \frac{1}{2}$ where

$$x_2 := \begin{pmatrix} 0 \\ 0 \\ \frac{1}{2} \end{pmatrix}.$$

We shall therefore try to prove a result similar to Theorem 1.3 for the RIP_L only in the case that $(\mathbf{s}, \mathbf{M})$ covers U. We need one further definition to state a result equivalent to Theorem 1.3 for the RIP_L. In equation (1.4) the bound on $||x - \widetilde{x}||_1$ involves $\sqrt{s}$. This arises because s is the maximum number of non-zero components that could be contained in an s-sparse vector. The equivalent for $(\mathbf{s}, \mathbf{M})$-sparse vectors is the following:

Definition 4.3. *The* number of elements *of a sparsity pattern* $(\mathbf{s}, \mathbf{M})$*, which we denote by* $\widetilde{s}$*, is given by*

$$\tilde{s} := s_1 + s_2 + \cdots + s_l$$

To prove that a sufficiently small RIP in levels constant implies an equation of the form (3.2), it is natural to adapt the steps used in [3] to prove Theorem 1.3. The adaptation of Theorem 1.3 yields a sufficient condition for recovery even in the noisy case.

Theorem 4.4. *Let* $(\mathbf{s}, \mathbf{M})$ *be a sparsity pattern with* l *levels and ratio constant* $\eta_{\mathbf{s},\mathbf{M}}$*. Suppose that the matrix* $U \in \mathbb{C}^{m \times n}$ *is covered by* $(\mathbf{s}, \mathbf{M})$ *and has a* RIP_L *constant* $\delta_{2\mathbf{s},\mathbf{M}}$ *satisfying*

$$\delta_{2\mathbf{s},\mathbf{M}} < \frac{1}{\sqrt{l\left(\sqrt{\eta_{\mathbf{s},\mathbf{M}}} + \frac{1}{4}\right)^2 + 1}}.$$

Let $x \in \mathbb{C}^n$ *and* $y \in \mathbb{C}^m$ *such that* $\|Ux - y\|_2 \leq \epsilon$*. Then any* $\widetilde{x} \in \mathbb{C}^n$ *which solve the* ℓ^1 *minimization problem*

$$\min_{\widehat{x} \in \mathbb{C}^n} \|\widehat{x}\|_1 \text{ subject to } \|U\widehat{x} - y\|_2 \leq \epsilon$$

also satisfy

$$\|x - \widetilde{x}\|_1 \leq C_1 \sigma_{\mathbf{s},\mathbf{M}}(x)_1 + D_1 \sqrt{\tilde{s}}\,\epsilon \tag{4.1}$$

and

$$\|x - \widetilde{x}\|_2 \leq \frac{\sigma_{\mathbf{s},\mathbf{M}}(x)_1}{\sqrt{\tilde{s}}} \left(C_2 + C_2' \sqrt[4]{l\eta_{\mathbf{s},\mathbf{M}}} \right) + \epsilon \left(D_2 + D_2' \sqrt[4]{l\eta_{\mathbf{s},\mathbf{M}}} \right) \tag{4.2}$$

where

$$\sigma_{\mathbf{s},\mathbf{M}}(x)_1 := \min \|x - \widehat{x_2}\|_1 \text{ such that } \widehat{x_2} \text{ is } (\mathbf{s},\mathbf{M})\text{-sparse}$$

and C_1, C_2, C_2', D_1, D_2 *and* D_2' *depend only on* $\delta_{2\mathbf{s},\mathbf{M}}$.

This result allows uniform recovery within the class of $(\mathbf{s},\mathbf{M})$-sparse vectors but the requirement on $\delta_{2\mathbf{s},\mathbf{M}}$ depends on l and $\eta_{\mathbf{s},\mathbf{M}}$. We make the following observations:

1. If we pick a sparsity pattern that uses lots of levels then we will require a better RIP_L constant.
2. If we pick a sparsity pattern with fewer levels then typically we shall pick a collection of s_i so that $\frac{s_i}{s_j}$ is correspondingly larger for distinct i and j.
3. If the RIP_L constant $\delta_{2\mathbf{s},\mathbf{M}}$ is sufficiently small so that the conclusion of Theorem 4.4, the bound on $\|x - x\|_2$ is weaker than the bound in Theorem 1.3.

As a consequence of these observations, at first glance it may appear that the results we have obtained with the RIP_L are weaker than those obtained using the standard RIP. However, Theorem 4.4 is stronger than Theorem 1.3 in two senses. Firstly, if one considers a sparsity pattern with one level then Theorem 4.4 reduces to Theorem 1.3. Secondly, the conclusion of Theorem 1.3 does not apply at all if we do not have the RIP. Therefore, for the examples given in Figure 2, Theorem 1.3 does not apply at all.

Ideally, it would be possible to find a constant C such that if the RIP_L constant is smaller than C then recovery of all $(\mathbf{s},\mathbf{M})$-sparse vectors would be possible. Unfortunately, we shall demonstrate that this is impossible in Theorems 4.5 and 4.6. Indeed, the result that we have proven is optimal up to constant terms, as the following results confirm.

Theorem 4.5. *Fix* $a \in \mathbb{N}$. *There exists* $m, n \in \mathbb{N}$, *a matrix* $U \in \mathbb{C}^{m \times n}$ *with two levels and a sparsity pattern* $(\mathbf{s},\mathbf{M})$ *that covers* U *such that the* RIP_L *constant* $\delta_{a\mathbf{s},\mathbf{M}}$ *and ratio constant* $\eta_{\mathbf{s},\mathbf{M}}$ *corresponding to* U *satisfy*

$$\delta_{a\mathbf{s},\mathbf{M}} \leq \frac{1}{|f(\eta_{\mathbf{s},\mathbf{M}})|}$$

where $f(\eta_{\mathbf{s},\mathbf{M}}) = o(\eta_{\mathbf{s},\mathbf{M}}^{\frac{1}{2}})$, *but there is an* $(\mathbf{s},\mathbf{M})$ *sparse* z^1 *such that*

$$z^1 \notin \arg\min \|z\|_1 \text{ such that } Uz = Uz^1.$$

Roughly speaking, Theorem 4.5 says that if we fix the number of levels and try to replace the condition

$$\delta_{2\mathbf{s},\mathbf{M}} < \frac{1}{\sqrt{l\left(\sqrt{\eta_{\mathbf{s},\mathbf{M}}} + \frac{1}{4}\right)^2 + 1}}$$

with a condition of the form

$$\delta_{2\mathbf{s},\mathbf{M}} < \frac{1}{C\sqrt{l}} \left(\eta_{\mathbf{s},\mathbf{M}}\right)^{-\frac{\alpha}{2}}$$

for some constant C and some $\alpha < 1$ then the conclusion of Theorem 4.4 ceases to hold. In particular, the requirement on $\delta_{2\mathbf{s},\mathbf{M}}$ cannot be independent of $\eta_{\mathbf{s},\mathbf{M}}$. The parameter a in the statement of Theorem 4.5 says that we cannot simply fix the issue by changing $\delta_{2\mathbf{s},\mathbf{M}}$ to $\delta_{3\mathbf{s},\mathbf{M}}$ or any further multiple of $\mathbf{s}$.

Similarly, we can state and prove a similar theorem that shows that the dependence on the number of levels, l, cannot be ignored.

Theorem 4.6. *Fix $a \in \mathbb{N}$. There exists $m, n \in \mathbb{N}$, a matrix $U \in \mathbb{C}^{m\times n}$ and a sparsity pattern $(\mathbf{s}, \mathbf{M})$ that covers U with ratio constant $\eta_{\mathbf{s},\mathbf{M}} = 1$ and l levels such that the RIP_L constant $\delta_{a\mathbf{s},\mathbf{M}}$ corresponding to U satisfies*

$$\delta_{a\mathbf{s},\mathbf{M}} \leq \frac{1}{|f(l)|}$$

where $f(l) = o(l^{\frac{1}{2}})$, but there is an $(\mathbf{s}, \mathbf{M})$ sparse z^1 such that

$$z^1 \notin \arg\min \|z\|_1 \text{ such that } Uz = Uz^1.$$

Furthermore, Theorem 4.7 shows that the ℓ^2 error estimate on $\|x - \widetilde{x}\|_2$ is optimal up to constant terms.

Theorem 4.7. *The ℓ^2 result (4.2) in Theorem 4.4 is sharp in the following sense:*

1. *For any $f(\eta) = o(\eta^{\frac{1}{4}})$, $g(\eta) = O(\sqrt{\eta})$ and $a \in \mathbb{N}$, there are natural numbers m and n, a matrix $U_2 \in \mathbb{C}^{m\times n}$ and a sparsity pattern $(\mathbf{s}, \mathbf{M})$ with two levels that such that*
   - $(\mathbf{s}, \mathbf{M})$ *covers* U_2
   - *The RIP_L constant corresponding to the sparsity pattern $(a\mathbf{s}, \mathbf{M})$, denoted by $\delta_{a\mathbf{s},\mathbf{M}}$, satisfies*
   $$\delta_{a\mathbf{s},\mathbf{M}} \leq \frac{1}{g(\eta_{\mathbf{s},\mathbf{M}})}.$$
   - *There exist vectors z and z^1 such that $U(z - z^1) = 0$ and $\|z\|_1 \leq \|z^1\|_1$ but*
   $$\|z - z^1\|_2 > \frac{f(\eta_{\mathbf{s},\mathbf{M}})}{\sqrt{\|\tilde{s}\|_1}} \sigma_{\mathbf{s},\mathbf{M}}(z^1)_1.$$
2. *For any $f(l) = o(l^{\frac{1}{4}})$, $g(l) = O(\sqrt{l})$ and $a \in \mathbb{N}$, there are natural numbers m and n, a matrix $U_2 \in \mathbb{C}^{m\times n}$ and a sparsity pattern $(\mathbf{s}, \mathbf{M})$ with $\eta_{\mathbf{s},\mathbf{M}} = 1$ such that*
   - $(\mathbf{s}, \mathbf{M})$ *covers* U_2
   - *The RIP_L constant corresponding to the sparsity pattern $(a\mathbf{s}, \mathbf{M})$, denoted by $\delta_{a\mathbf{s},\mathbf{M}}$, satisfies*
   $$\delta_{a\mathbf{s},\mathbf{M}} \leq \frac{1}{g(l)}.$$
   - *There exist vectors z and z^1 such that $U(z - z^1) = 0$ and $\|z\|_1 \leq \|z^1\|_1$ but*
   $$\|z - z^1\|_2 > \frac{f(l)}{\sqrt{\|\tilde{s}\|_1}} \sigma_{\mathbf{s},\mathbf{M}}(z^1)_1.$$

Theorems of a similar form to Theorem 1.3 typically begin by proving the *robust ℓ^2 nullspace property of order s*.

Definition 4.8. *A matrix $U \in \mathbb{C}^{m\times n}$ is said to satisfy the ℓ^2 robust nullspace property of order s if there is a $\rho \in (0, 1)$ and a $\tau > 0$ such that for all vectors $v \in \mathbb{C}^n$ and all S which are subsets of $\{1, 2, 3, \dots, n\}$ with $|S| \leq s$,*

$$\|v_S\|_2 \leq \frac{\rho}{\sqrt{s}} \|v_{S^c}\|_1 + \tau \|Uv\|_2.$$

The implication is then the following Theorem: (for example, see [18], Theorem 4.22)

Theorem 4.9. *Suppose that $U \in \mathbb{C}^{m\times n}$ satisfies the ℓ^2 robust nullspace property of order s with constants $\rho \in (0, 1)$ and $\tau > 0$. Then any solutions $\widetilde{x} \in \mathbb{C}^n$ to the ℓ^1 minimization problem*

$$\min_{\widehat{x}\in\mathbb{C}^n} \|\widehat{x}\|_1 \text{ subject to } \|U\widehat{x} - y\|_2 \leq \epsilon$$

where $\|Ux - y\|_2 \leq \epsilon$ satisfy

$$\|\widetilde{x} - x\|_1 \leq C_1 \sigma_{\mathbf{s},\mathbf{M}}(x) + D_1 \epsilon \sqrt{s}$$
$$\|\widetilde{x} - x\|_2 \leq \frac{C_2 \sigma_{\mathbf{s},\mathbf{M}}(x)}{\sqrt{s}} + D_2 \epsilon$$

where the constants C_1, C_2, D_1 and D_2 depend only on ρ and τ.

The corresponding natural extension of the ℓ^2 robust nullspace of order s to the $(\mathbf{s},\mathbf{M})$-sparse case is the ℓ^2 *robust nullspace property of order* $(\mathbf{s},\mathbf{M})$.

Definition 4.10. *A matrix $U \in \mathbb{C}^{m\times n}$ satisfies the ℓ^2 robust nullspace property of order $(\mathbf{s},\mathbf{M})$ if there is a $\rho \in (0,1)$ and a $\tau > 0$ such that*

$$\|v_S\|_2 \leq \frac{\rho}{\sqrt{\tilde{s}}}\|v_{S^c}\|_1 + \tau\|Uv\|_2 \tag{4.3}$$

for all $(\mathbf{s},\mathbf{M})$-sparse sets S and vectors $v \in \mathbb{C}^n$.

The $(\mathbf{s},\mathbf{M})$-sparse version of Theorem 4.9 is Theorem 4.11.

Theorem 4.11. *Suppose that a matrix $U \in \mathbb{C}^{m\times n}$ satisfies the ℓ^2 robust nullspace property of order $(\mathbf{s},\mathbf{M})$ with constants $\rho \in (0,1)$ and $\tau > 0$. Let $x \in \mathbb{C}^n$ and $y \in \mathbb{C}^m$ such that $\|Ux - y\|_2 < \epsilon$. Then any solutions $\widetilde{x}$ of the ℓ^1 minimization problem*

$$\min_{\widehat{x}\in\mathbb{C}^n} \|\widehat{x}\|_1 \text{ subject to } \|U\widehat{x} - y\|_2 \leq \epsilon$$

satisfy

$$\|\widetilde{x} - x\|_1 \leq A_1\sigma_{\mathbf{s},\mathbf{M}}(x)_1 + C_1\epsilon\sqrt{\tilde{s}} \tag{4.4}$$

$$\|\widetilde{x} - x\|_2 \leq \frac{\sigma_{\mathbf{s},\mathbf{M}}(x)_1}{\sqrt{\tilde{s}}}\left(A_2 + B_2\sqrt[4]{l\eta_{\mathbf{s},\mathbf{M}}}\right) + 2\epsilon\left(C_2 + D_2\sqrt[4]{l\eta_{\mathbf{s},\mathbf{M}}}\right) \tag{4.5}$$

where

$$A_1 := \frac{2+2\rho}{1-\rho}, \quad C_1 := \frac{4\tau}{1-\rho}, \quad A_2 := \frac{2\rho+2\rho^2}{1-\rho},$$

$$B_2 := \frac{\left(2\sqrt{\rho}+1\right)(1+\rho)}{1-\rho}, \quad C_2 := \frac{\rho\tau+\tau}{1-\rho} \ \textit{and} \ D_2 := \frac{4\sqrt{\rho}\tau + 3\tau - \rho\tau}{2-2\rho}.$$

This Theorem explains where the dependence on $\eta_{\mathbf{s},\mathbf{M}}$ and l in the estimate $\|x - \widetilde{x}\|_2$ in Theorem 4.4 emerges from. Analogous to proving the ℓ^2 error estimates in Theorem 1.3 using the ℓ^2 robust nullspace property of order s, we prove the ℓ^2 error estimate in Theorem 4.4 by showing that a sufficiently small RIP_L constant implies the robust ℓ^2 nullspace property of order $(\mathbf{s},\mathbf{M})$. The ℓ^2 error estimate (4.5) follows and we are left with a dependence on $\sqrt[4]{l\eta_{\mathbf{s},\mathbf{M}}}$ in the right hand side of (4.2). As before, Theorem 4.12 shows that this is optimal.

Theorem 4.12. *The result in Theorem 4.11 is sharp, in the sense that*

1. *For any f satisfying*

   $$f(\rho,\tau,\eta) = o(\eta^{\frac{1}{4}}) \textit{ for fixed } \rho \in (0,1) \textit{ and } \tau > 0$$

   there exists natural numbers m and n, a matrix $U_2 \in \mathbb{C}^{m\times n}$ and a sparsity pattern $(\mathbf{s},\mathbf{M})$ with ratio constant $\eta_{\mathbf{s},\mathbf{M}}$ and two levels such that

   - *$(\mathbf{s},\mathbf{M})$ covers U_2*
   - *U_2 satisfies the ℓ^2 robust nullspace property of order $(\mathbf{s},\mathbf{M})$ with constants $\rho \in (0,1)$ and $\tau > 0$*
   - *There exist vectors z and z^1 such that $U_2(z - z^1) = 0$ and $\|z\|_1 \leq \|z^1\|_1$ but*

   $$\|z - z^1\|_2 > \frac{f(\rho,\tau,\eta_{\mathbf{s},\mathbf{M}})}{\sqrt{\tilde{s}}}\sigma_{\mathbf{s},\mathbf{M}}(z^1)_1.$$

2. *For any f satisfying*

   $$f(\rho,\tau,l) = o(l^{\frac{1}{4}}) \textit{ for fixed } \rho \in (0,1) \textit{ and } \tau > 0$$

   there exists natural numbers m, n, a matrix $U_2 \in \mathbb{C}^{m\times n}$ and a sparsity pattern $(\mathbf{s},\mathbf{M})$ with ratio constant $\eta_{\mathbf{s},\mathbf{M}} = 1$ and l levels such that

   - *$(\mathbf{s},\mathbf{M})$ covers U_2*
   - *U_2 satisfies the ℓ^2 robust nullspace property of order $(\mathbf{s},\mathbf{M})$ with constants $\rho \in (0,1)$ and $\tau > 0$*

- *There exist vectors z and z^1 such that $U(z - z^1) = 0$ and $\|z\|_1 \leq \|z^1\|_1$ but*
$$\|z - z^1\|_2 > \frac{f(\rho, \tau, l)}{\sqrt{\tilde{s}}} \sigma_{\mathbf{s},\mathbf{M}}(z^1)_1.$$

The conclusions that we can draw from the above theorems are the following:

1. The RIP_L will guarantee ℓ^1 and ℓ^2 estimates on the size of the error $\|\widetilde{x} - x\|$, provided that the RIP_L constant is sufficiently small (Theorem 4.4).
2. The requirement that the RIP_L constant is sufficiently small is dependent on $\sqrt{\eta_{\mathbf{s},\mathbf{M}}}$ and $\sqrt{l}$. This is optimal up to constants (Theorem 4.5 and Theorem 4.6).
3. The ℓ^2 error when using the RIP_L has additional factors of the form $\sqrt[4]{l}$ and $\sqrt[4]{\eta_{\mathbf{s},\mathbf{M}}}$. Again, these are optimal up to constants (Theorem 4.7).
4. Typically, we use the robust ℓ^2 nullspace property of order s to give us a bound on the ℓ^2 error in using $\widetilde{x}$ to approximate x. When this concept is extended to a robust ℓ^2 nullspace property of order $(\mathbf{s}, \mathbf{M})$ then the ℓ^2 error gains additional factors of the form $\sqrt[4]{l}$ and $\sqrt[4]{\eta_{\mathbf{s},\mathbf{M}}}$ (Theorem 4.11).
5. These factors are optimal up to constants, so that even if we ignore the RIP_L and still try to prove results using the ℓ^2 robust nullspace property of order $(\mathbf{s}, \mathbf{M})$ then we would be unable to improve the ℓ^2 error (Theorem 4.12).

We have demonstrated that the RIP in levels may be able to explain why permutations within levels are possible and why more general permutations are impossible. The results that we have obtained give a sufficient condition on the RIP in levels constant that guarantees $(\mathbf{s}, \mathbf{M})$-sparse recovery. Furthermore, we have managed to demonstrate that this condition and the conclusions that follow from it are optimal up to constants.

5 Conclusions and open problems

The flip test demonstrates that in practical applications the ability to recover sparse signals depends on the structure of the sparsity, so that a tool that guarantees uniform recovery of all s-sparse signals does not apply. The flip test with permutations within the levels suggests that reasonable sampling schemes provide a different form of uniform recovery, namely, the recovery of $(\mathbf{s}, \mathbf{M})$-sparse signals. It is therefore natural to try to find theoretical tools that are able to analyse and describe this phenomenon. However, we are now left with the fundamental problems:

- Given a sampling basis (say Fourier) and a recovery basis (say wavelets) and a sparsity pattern $(\mathbf{s}, \mathbf{M})$, what kind of sampling procedure will give the RIP_L according to $(\mathbf{s}, \mathbf{M})$?
- How many samples must one draw, and how should they be distributed (as a function of $(\mathbf{s}, \mathbf{M})$), in order to obtain the RIP_L?

Note that these problems are vast as the sampling patterns will not only depend on the sparsity patterns, but of course also on the sampling basis and recovery basis (or frame). Thus, covering all interesting cases in application will yield an incredibly rich mathematical theory.

Acknowledgements

The authors would like to thank Arash Amini for asking an excellent question, during a talk given at EPFL, about the possibility of having a theory for the RIP in levels. It was this question that sparked the research leading to this paper. The authors would also like to thank Ben Adcock for many great suggestions and input.

ACH acknowledges support from a Royal Society University Research Fellowship as well as the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/L003457/1

6 Proofs

We shall present the proofs in a different arrangement to the order in which their statements were presented. The first proof that we shall present is that of Theorem 4.11.

6.1 Proof of Theorem 4.11

We begin with the following lemma:

Lemma 6.1. *Suppose that $U \in \mathbb{C}^{m\times n}$ satisfies the ℓ^2 robust nullspace property of order $(\mathbf{s},\mathbf{M})$ with constants $\rho \in (0,1)$ and $\tau > 0$. Fix $v \in \mathbb{C}^n$, and let S be an $(\mathbf{s},\mathbf{M})$-sparse set such that.*

1. *For all $(\mathbf{s},\mathbf{M})$-sparse sets T, $\|v_S\|_1 \geq \|v_T\|_1$.*
2. *$|S| = \tilde{s}$.*

Then

$$\|v\|_2 \leq \frac{\|v_{S^c}\|_1}{\sqrt{\tilde{s}}}\left[\rho + \left(\sqrt{\rho}+\frac{1}{2}\right)\sqrt[4]{l\eta_{\mathbf{s},\mathbf{M}}}\right] + \tau\|Uv\|_2\left[\frac{\sqrt[4]{l\eta_{\mathbf{s},\mathbf{M}}}}{2}+1\right].$$

Proof. For $i = 1,2,\dots,l$, we define S_0^i to be $S_0^i = S \cap \{M_{i-1}+1, M_{i-1}+2,\dots,M_i\}$. Let

$$m = \max_{i=1,2,\dots,l}\min_{j\in S_0^i}|v_j|.$$

Since $|S_0^i| = s_i$ (otherwise $|S| < \tilde{s}$), we can see that given any $i = 1,2,\dots,l$

$$\|v_S\|_2 = \sqrt{\sum_{n\in S}|v|_n^2} \geq \sqrt{\sum_{j\in S_0^i}|v|_j^2} \geq \sqrt{s_i}\min_{j\in S_0^i}|v_j| \geq \min_{k=1,2,\dots,l}\sqrt{s_k}\min_{j\in S_0^i}|v_j|$$

so that $\|v_S\|_2 \geq m \min_{k=1,2\dots,l}\sqrt{s_k}$. Furthermore, $|v_j| \leq m$ for each $j \in S^c$ otherwise there is an $(\mathbf{s},\mathbf{M})$-sparse T with $\|v_T\|_1 > \|v_S\|_1$.

Therefore

$$\|v_{S^c}\|_2^2 = \sum_{j\in S^c}|v_j|^2 \leq \sum_{j\in S^c} m|v_j| \leq \frac{\|v_{S^c}\|_1\|v_S\|_2}{\min\sqrt{s_i}}.$$

By the ℓ^2 robust nullspace property of order $(\mathbf{s},\mathbf{M})$,

$$\|v_{S^c}\|_1\|v_S\|_2 \leq \frac{\rho}{\sqrt{\tilde{s}}}\|v_{S^c}\|_1^2 + \tau\|Uv\|_2\|v_{S^c}\|_1.$$

Since $\sqrt{a+b} \leq \sqrt{a}+\sqrt{b}$,

$$\|v_{S^c}\|_2 \leq \frac{1}{\min\sqrt[4]{s_i}}\left(\frac{\sqrt{\rho}}{\sqrt[4]{\tilde{s}}}\|v_{S^c}\|_1 + \sqrt{\tau\|Uv\|_2\|v_S^c\|_1}\right) \tag{6.1}$$

Using the arithmetic-geometric mean inequality,

$$\sqrt{\tau\|Uv\|_2\|v_{S^c}\|_1} = \sqrt{\tau\|Uv\|_2\sqrt[4]{\tilde{s}}\frac{\|v_{S^c}\|_1}{\sqrt[4]{\tilde{s}}}} \leq \frac{\tau\|Uv\|_2\sqrt[4]{\tilde{s}}}{2} + \frac{\|v_{S^c}\|_1}{2\sqrt[4]{\tilde{s}}}$$

Therefore, (6.1) yields

$$\begin{aligned}\|v_{S^c}\|_2 &\leq \frac{1}{\min\sqrt[4]{s_i}}\left(\frac{\sqrt{\rho}}{\sqrt[4]{\tilde{s}}}\|v_{S^c}\|_1 + \frac{\|v_{S^c}\|_1}{2\sqrt[4]{\tilde{s}}} + \frac{\tau\|Uv\|_2\sqrt[4]{\tilde{s}}}{2}\right)\\ &\leq \frac{\|v_{S^c}\|_1}{\sqrt[4]{\tilde{s}}\min\sqrt[4]{s_i}}\left(\sqrt{\rho}+\frac{1}{2}\right) + \frac{\tau\|Uv\|_2\sqrt[4]{l\eta_{\mathbf{s},\mathbf{M}}}}{2}\end{aligned}$$

because $\frac{\tilde{s}}{\min s_i} \leq l\eta_{\mathbf{s},\mathbf{M}}$. Once again, employing the ℓ^2 nullspace property gives

$$\begin{aligned}\|v\|_2 &\leq \|v_S\|_2 + \|v_{S^c}\|_2\\ &\leq \frac{\rho}{\sqrt{\tilde{s}}}\|v_{S^c}\|_1 + \tau\|Uv\|_2 + \frac{\|v_{S^c}\|_1}{\sqrt[4]{\tilde{s}}\min\sqrt[4]{s_i}}\left(\sqrt{\rho}+\frac{1}{2}\right) + \frac{\tau\|Uv\|_2\sqrt[4]{l\eta_{\mathbf{s},\mathbf{M}}}}{2}\\ &\leq \frac{\|v_{S^c}\|_1}{\sqrt{\tilde{s}}}\left[\rho + \left(\sqrt{\rho}+\frac{1}{2}\right)\frac{\sqrt[4]{\tilde{s}}}{\min\sqrt[4]{s_i}}\right] + \tau\|Uv\|_2\left[\frac{\sqrt[4]{l\eta_{\mathbf{s},\mathbf{M}}}}{2}+1\right]\\ &\leq \frac{\|v_{S^c}\|_1}{\sqrt{\tilde{s}}}\left[\rho + \left(\sqrt{\rho}+\frac{1}{2}\right)\sqrt[4]{l\eta_{\mathbf{s},\mathbf{M}}}\right] + \tau\|Uv\|_2\left[\frac{\sqrt[4]{l\eta_{\mathbf{s},\mathbf{M}}}}{2}+1\right].\end{aligned}$$

□

The remaining error estimates will follow from various properties related to the ℓ^1 *robust nullspace property* (see [18], definition 4.17) holds. To be precise,

Definition 6.2. *A matrix $U \in \mathbb{C}^{m\times n}$ satisfies the ℓ^1 robust null space property relative to S with constants $\rho \in (0,1)$ and $\tau' > 0$ if*

$$\|v_S\|_1 \leq \rho\|v_{S^c}\|_1 + \tau'\|Uv\|_2 \tag{6.2}$$

for any $v \in \mathbb{C}^n$. We say that U satisfies the ℓ^1 robust null space property of order $(\mathbf{s},\mathbf{M})$ if (6.2) holds for any $(\mathbf{s},\mathbf{M})$-sparse sets S.

It is easy to see that if U satisfies the ℓ^2 robust nullspace property of order $(\mathbf{s},\mathbf{M})$ with constants ρ and τ then, for any $(\mathbf{s},\mathbf{M})$-sparse set S, U also satisfies the ℓ^1 robust nullspace property relative to S with constants ρ and $\tau\sqrt{\tilde{s}}$. Indeed, assume that U satisfies the ℓ^2 robust nullspace property of order $(\mathbf{s},\mathbf{M})$ with constants ρ and τ. Then

$$\|v_S\|_1 \leq \sqrt{\tilde{s}}\|v_S\|_2 \leq \rho\|v_{S^c}\|_1 + \tau\sqrt{\tilde{s}}\|Uv\|_2.$$

An immediate conclusion of the robust null space property is the following, proven in [18] as Theorem 4.20.

Lemma 6.3. *Suppose that $U \in \mathbb{C}^{m\times n}$ satisfies the ℓ^1 robust null space property with constants $\rho \in (0,1)$ and τ' relative to a set S. Then for any complex vectors $x, z \in \mathbb{C}^n$,*

$$\|z-x\|_1 \leq \frac{1+\rho}{1-\rho}\left(\|z\|_1 - \|x\|_1 + 2\|x_{S^c}\|_1\right) + \frac{2\tau'}{1-\rho}\|U(z-x)\|_2.$$

We can use this lemma to show the following important result, which is similar both in proof and statement to Theorem 4.19 in [18].

Lemma 6.4. *Suppose that a matrix $U \in \mathbb{C}^{m\times n}$ satisfies the ℓ^1 robust nullspace property of order $(\mathbf{s},\mathbf{M})$ with constants $\rho \in (0,1)$ and $\tau' > 0$. Furthermore, suppose that $\|Ux-y\|_2 \leq \epsilon$. Then any solutions $\widetilde{x}$ to the ℓ^1 minimization problem*

$$\min_{\widehat{x}\in\mathbb{C}^n} \|\widehat{x}\|_1 \text{ subject to } \|U\widehat{x}-y\|_2 \leq \epsilon$$

satisfy

$$\|x-\widetilde{x}\|_1 \leq \frac{2+2\rho}{1-\rho}\sigma_{\mathbf{s},\mathbf{M}}(x)_1 + \frac{4\tau'\epsilon}{1-\rho}.$$

Proof. By Lemma 6.3, for any $(\mathbf{s},\mathbf{M})$-sparse set S

$$\|x-\widetilde{x}\|_1 \leq \frac{1+\rho}{1-\rho}\left(\|\widetilde{x}\|_1 - \|x\|_1 + 2\|x_{S^c}\|_1\right) + \frac{2\tau'}{1-\rho}\|U(x-\widetilde{x})\|_2$$

Because both $\|Ux-y\|_2$ and $\|U\widetilde{x}-y\|_2$ are smaller than or equal to ϵ, $\|Ux-U\widetilde{x}\| \leq 2\epsilon$. Furthermore, because $\widetilde{x}$ has minimal ℓ_1 norm, $\|\widetilde{x}\|_1 - \|x\|_1 \leq 0$. Thus

$$\|x-\widetilde{x}\|_1 \leq \frac{2+2\rho}{1-\rho}\|x_{S^c}\|_1 + \frac{4\tau'\epsilon}{1-\rho}.$$

If we take S to be the $(\mathbf{s},\mathbf{M})$-sparse set which maximizes $\|x_S\|_1$, then

$$\|x-\widetilde{x}\|_1 \leq \frac{2+2\rho}{1-\rho}\sigma_{\mathbf{s},\mathbf{M}}(x)_1 + \frac{4\tau'\epsilon}{1-\rho}.$$

□

We can combine these results to complete the proof of Theorem 4.11. Indeed, (4.4) follows immediately from Lemma 6.4 and the fact that U satisfies the ℓ^1 robust nullspace property with constants ρ and $\tau\sqrt{\|\mathbf{s}\|_1}$. To prove (4.5), we can simply set $v = x - \widetilde{x}$ in Lemma 6.1 to see that

$$\begin{aligned}\|x-\widetilde{x}\|_2 &\leq \frac{\|(x-\widetilde{x})_{S^c}\|_1}{\sqrt{\tilde{s}}}\left[\rho + \left(\sqrt{\rho}+\frac{1}{2}\right)\sqrt[4]{l\eta_{\mathbf{s},\mathbf{M}}}\right] + \tau\|U\left(x-\widetilde{x}\right)\|_2\left[\frac{\sqrt[4]{l\eta_{\mathbf{s},\mathbf{M}}}}{2}+1\right] \\ &\leq \frac{\|x-\widetilde{x}\|_1}{\sqrt{\tilde{s}}}\left[\rho + \left(\sqrt{\rho}+\frac{1}{2}\right)\sqrt[4]{l\eta_{\mathbf{s},\mathbf{M}}}\right] + 2\tau\epsilon\left[\frac{\sqrt[4]{l\eta_{\mathbf{s},\mathbf{M}}}}{2}+1\right]\end{aligned}$$

and the result follows from (4.4).

6.2 Proof of Theorem 4.4

It will suffice to prove that the conditions on U and $(\mathbf{s}, \mathbf{M})$ in Theorem 4.4 imply the ℓ^2 robust nullspace property. To show this, we begin by stating the following inequality, proven in [4]

Lemma 6.5 (The norm inequality for ℓ_1 and ℓ_2)**.** *Let $v = (v_1, v_2, \ldots, v_s)$ where $v_1 \geq v_2 \geq v_3 \geq \cdots \geq v_s$. Then*

$$\|v\|_2 \leq \frac{1}{\sqrt{s}}\|v\|_1 + \frac{\sqrt{s}}{4}(v_1 - v_s)$$

We will now prove the following additional lemma which is almost identical in statement and proof to that of Lemma 6.1 in [3].

Lemma 6.6. *Suppose that $x, y \in \Sigma_{\mathbf{s},\mathbf{M}}$ and that*

$$\|Ux\|_2^2 - \|x\|_2^2 = t\|x\|_2^2. \tag{6.3}$$

Furthermore, suppose that x and y are orthogonal. Then $|\langle Ux, Uy\rangle| \leq \sqrt{\delta_{2\mathbf{s},\mathbf{M}}^2 - t^2}\|x\|_2\|y\|_2$ where $\delta_{2\mathbf{s},\mathbf{M}}$ is the restricted isometry constant corresponding to the sparsity pattern $(2\mathbf{s}, \mathbf{M})$ and the matrix U.

Proof. Without loss of generality, we can assume that $\|x\|_2 = \|y\|_2 = 1$. Note that for $\alpha, \beta \in \mathbb{R}$ and $\gamma \in \mathbb{C}$, the vectors $\alpha x + \gamma y$ and $\beta x - \gamma y$ are contained in $\Sigma_{2\mathbf{s},\mathbf{M}}$. Therefore,

$$\begin{aligned}\|U(\alpha x + \gamma y)\|_2^2 &\leq (1 + \delta_{2\mathbf{s},\mathbf{M}})\|\alpha x + \gamma y\|_2^2 \\ &= (1 + \delta_{2\mathbf{s},\mathbf{M}})\left(\langle \alpha x, \alpha x\rangle + \langle \gamma y, \alpha x\rangle + \langle \alpha x, \gamma y\rangle + \langle \gamma y, \gamma y\rangle\right) \\ &= (1 + \delta_{2\mathbf{s},\mathbf{M}})(\alpha^2 + |\gamma|^2)\end{aligned} \tag{6.4}$$

where the last line follows because $\langle x, y\rangle = 0$ (from the orthogonality of x and y). Similarly,

$$-\|U(\beta x - \gamma y)\|_2^2 \leq -(1 - \delta_{2\mathbf{s},\mathbf{M}})(\beta^2 + |\gamma|^2) \tag{6.5}$$

We will now add these two inequalities. On the one hand (by using the assumption in (6.3) and the fact that α,β are real), we have

$$\begin{aligned}\|U(\alpha x + \gamma y)\|_2^2 - \|U(\beta x - \gamma y)\|_2^2 &= \alpha^2\|Ux\|_2^2 + 2\text{Re}(\alpha\overline{\gamma}\langle Ux, Uy\rangle) + |\gamma|^2\|Uy\|_2^2 \\ &\quad - \left(\beta^2\|Ux\|_2^2 - 2\text{Re}(\beta\overline{\gamma}\langle Ux, Uy\rangle) + |\gamma|^2\|Uy\|_2^2\right) \\ &= (1+t)\left(\alpha^2 - \beta^2\right) + 2(\alpha+\beta)\text{Re}(\overline{\gamma}\langle Ux, Uy\rangle)\end{aligned}$$

and on the other hand (from (6.4) and (6.5))

$$\|U(\alpha x + \gamma y)\|_2^2 - \|U(\beta x - \gamma y)\|_2^2 \leq \delta_{2\mathbf{s},\mathbf{M}}\left(\alpha^2 + \beta^2 + 2|\gamma|^2\right) + \alpha^2 - \beta^2.$$

Therefore,

$$(1+t)\left(\alpha^2 - \beta^2\right) + 2(\alpha+\beta)\text{Re}(\overline{\gamma}\langle Ux, Uy\rangle) \leq \delta_{2\mathbf{s},\mathbf{M}}\left(\alpha^2 + \beta^2 + 2|\gamma|^2\right) + \alpha^2 - \beta^2.$$

After choosing γ so that $\text{Re}(\overline{\gamma}\langle Ux, Uy\rangle) = |\langle Ux, Uy\rangle|$ we obtain

$$|\langle Ux, Uy\rangle| \leq \frac{1}{2\alpha + 2\beta}\left((\delta_{2\mathbf{s},\mathbf{M}} - t)\alpha^2 + (\delta_{2\mathbf{s},\mathbf{M}} + t)\beta^2 + 2\delta_{2\mathbf{s},\mathbf{M}}\right) \tag{6.6}$$

because $|\gamma| = 1$. By the definition of the RIP in levels constant, $\delta_{2\mathbf{s},\mathbf{M}} \geq \delta_{\mathbf{s},\mathbf{M}}$ and so

$$|t| = \left|\|Ux\|_2^2 - \|x\|_2^2\right| \leq \delta_{\mathbf{s},\mathbf{M}} \leq \delta_{2\mathbf{s},\mathbf{M}}.$$

Therefore, $\sqrt{\frac{\delta_{2\mathbf{s},\mathbf{M}}+t}{\delta_{2\mathbf{s},\mathbf{M}}-t}} \in \mathbb{R}$ and so we can set $\alpha = \sqrt{\frac{\delta_{2\mathbf{s},\mathbf{M}}+t}{\delta_{2\mathbf{s},\mathbf{M}}-t}}$ and $\beta = \frac{1}{\alpha}$ in equation (6.6). With these values, we obtain

$$\begin{aligned}|\langle Ux, Uy\rangle| &\leq \frac{\alpha}{2\alpha^2 + 2}\left(\delta_{2\mathbf{s},\mathbf{M}} + t + \delta_{2\mathbf{s},\mathbf{M}} - t + 2\delta_{2\mathbf{s},\mathbf{M}}\right) \\ &\leq \frac{4\delta_{2\mathbf{s},\mathbf{M}}\alpha(\delta_{2\mathbf{s},\mathbf{M}} - t)}{4\delta_{2\mathbf{s},\mathbf{M}}} \leq \sqrt{\delta_{2\mathbf{s},\mathbf{M}}^2 - t^2}.\end{aligned}$$

□

Proof (of Theorem 4.4). Let $x \in \mathbb{C}^m$ be an arbitrary m dimensional complex vector, and let

$$x^i := x_{\{M_{i-1}+1, M_{i-1}+2, \dots, M_i\}}$$

denote the ith level of x. For an arbitrary vector $v = (v_1, v_2, \dots, v_n)$, we define $|v|$ to be the vector $(|v_1|, |v_2|, \dots, |v_n|)$. Furthermore, let S_0^i denote the indexes of the s_ith largest elements of $|x^i|$ (if there are less than s_i elements in x^i then we zero pad x^i until it has s_i elements in it), and $S_0 := \bigcup_{i=1}^l S_0^i$. We then define S_1^i to be the indexes of the s_ith largest elements of $|x^i|$ that are not contained in S_0^i (again, if there are less than s_i of these elements then we zero pad x^i until there are s_i elements in $x^i_{(S_0^i)^c}$) and define $S_1 := \bigcup_{i=1}^l S_1^i$. In general, we can make a similar definition to form a collection of index sets labelled $(S_j^i)_{i=1,2\dots,l,j=1,2,\dots}$ and corresponding $(\mathbf{s}, \mathbf{M})$-sparse S_j.

These definitions and the fact that $(\mathbf{s}, \mathbf{M})$ covers U implies that if $\Omega = \bigcup_{j \geq 0} S_j$ then

$$x_\Omega = x. \tag{6.7}$$

By the definition of S_0, $\|x_\Lambda\|_1 \leq \|x_{S_0}\|_1$ whenever Λ is $(\mathbf{s}, \mathbf{M})$-sparse. By Theorem 4.11, it will suffice to verify that

$$\sqrt{\|\mathbf{s}\|_1}\|x_{S_0}\|_2 \leq \rho \|x_{S_0^c}\|_1 + \tau \sqrt{\|\mathbf{s}\|_1}\|Ux\|_2 \tag{6.8}$$

holds. Set

$$\|Ux_{S_0}\|_2^2 = (1+t)\|x_{S_0}\|_2^2. \tag{6.9}$$

Clearly, $|t| \leq \delta_{\mathbf{s},\mathbf{M}}$. Then

$$\begin{aligned}
\|Ux_{S_0}\|_2^2 &= \langle Ux_{S_0}, Ux_{S_0} \rangle \\
&= \langle Ux_{S_0}, Ux \rangle - \sum_{j \geq 1} \left\langle Ux_{S_0}, Ux_{S_j} \right\rangle. \qquad (6.10)
\end{aligned}$$

where we have used (6.7). The first term on the right hand side can be bounded above by $\|Ux_{S_0}\|_2 \|Ux\|_2$ using the Cauchy-Schwarz inequality. Then (using (6.9))

$$|\langle Ux_{S_0}, Ux \rangle| \leq \sqrt{1+t}\|x_{S_0}\|_2 \|Ux\|_2. \tag{6.11}$$

Furthermore, we can use Lemma 6.6 to see that

$$\begin{aligned}
\left| \sum_{j \geq 1} \left\langle Ux_{S_0}, Ux_{S_j} \right\rangle \right| &\leq \sqrt{\delta_{2\mathbf{s}}^2 - t^2} \sum_{j \geq 1} \|x_{S_0}\|_2 \|x_{S_j}\|_2 \\
&\leq \|x_{S_0}\|_2 \sqrt{\delta_{2\mathbf{s}}^2 - t^2} \sum_{i=1}^l \sum_{j \geq 1} \|x_{S_j^i}\|_2 \qquad (6.12)
\end{aligned}$$

Let $x_{i,j}^+$ (correspondingly $x_{i,j}^-$) be the largest element of $\left|x_{S_j^i}\right|$ (correspondingly the smallest element of $\left|x_{S_j^i}\right|$). By the norm inequality for ℓ_1 and ℓ_2 (Lemma 6.5),

$$\begin{aligned}
\sum_{j \geq 1} \|x_{S_j^i}\|_2 &\leq \sum_{j \geq 1} \left(\frac{1}{\sqrt{s_i}} \|x_{S_j^i}\|_1 \right) + \frac{\sqrt{s_i}}{4} \sum_{j \geq 1} \left(x_{i,j}^+ - x_{i,j}^- \right) \\
&\leq \sum_{j \geq 1} \left(\frac{1}{\sqrt{s_i}} \|x_{S_j^i}\|_1 \right) + \frac{\sqrt{s_i}}{4} \left(x_{i,1}^+ + \sum_{j \geq 2} x_{i,j}^+ - \sum_{j \geq 1} x_{i,j}^+ \right) \\
&\leq \sum_{j \geq 1} \left(\frac{1}{\sqrt{s_i}} \|x_{S_j^i}\|_1 \right) + \frac{\sqrt{s_i}}{4} \left(x_{i,1}^+ + \sum_{j \geq 1} \left(x_{i,j+1}^+ - x_{i,j}^- \right) \right) \\
&\leq \sum_{j \geq 1} \left(\frac{1}{\sqrt{s_i}} \|x_{S_j^i}\|_1 \right) + \frac{\sqrt{s_i}}{4} x_{i,1}^+
\end{aligned}$$

where the last line follows because $x^+_{i,j+1} - x^-_{i,j} \leq 0$.
Additionally,

$$\frac{\sqrt{s_i}}{4} x^+_{i,1} = \frac{1}{4}\sqrt{\underbrace{(x^+_{i,1})^2 + (x^+_{i,1})^2 + (x^+_{i,1})^2 + \cdots + (x^+_{i,1})^2}_{s_i}} \leq \frac{1}{4}\|x_{S_0^i}\|_2$$

because each element of $\left|x_{S_0^i}\right|$ is larger than $x^+_{i,1}$. We conclude that

$$\begin{aligned}\sum_{i=1}^{l}\sum_{j\geq 1}\|x_{S_j^i}\|_2 &\leq \sum_{j\geq 1}\sum_{i=1}^{l}\frac{1}{\sqrt{s_i}}\|x_{S_j^i}\|_1 + \sum_{i=1}^{l}\frac{1}{4}\|x_{S_0^i}\|_2 \\ &\leq \frac{1}{\min\sqrt{s_i}}\sum_{j\geq 1}\sum_{i=1}^{l}\|x_{S_j^i}\|_1 + \frac{1}{4}\sqrt{l}\|x_{S_0}\|_2 \\ &\leq \frac{1}{\min\sqrt{s_i}}\sum_{j\geq 1}\|x_{S_j}\|_1 + \frac{1}{4}\sqrt{l}\|x_{S_0}\|_2 \\ &\leq \frac{1}{\min\sqrt{s_i}}\left\|x_{\bigcup\limits_{j\geq 1} S_j}\right\|_1 + \frac{1}{4}\sqrt{l}\|x_{S_0}\|_2\end{aligned}$$

where the second line follows from the Cauchy-Schwarz inequality applied to the vectors $(\underbrace{1,1,\ldots,1}_{l})$ and $(\|x_{S_0^1}\|_2, \|x_{S_0^2}\|_2, \ldots, \|x_{S_0^l}\|_2)$ and the third line and fourth line follows from the disjoint supports of the vectors $x_{S_j^i}$ and $x_{S_{j'}^{i'}}$ whenever $i \neq i'$ or $j \neq j'$. By (6.7) and the disjointedness of S_i, S_j for $i \neq j$, $\bigcup\limits_{j\geq 1} S_j = S_0^c$ so

$$\sum_{i=1}^{l}\sum_{j\geq 1}\|x_{S_j^i}\|_2 \leq \frac{1}{\min\sqrt{s_i}}\|x_{S_0^c}\|_1 + \frac{1}{4}\sqrt{l}\|x_{S_0}\|_2. \tag{6.13}$$

Combining (6.9),(6.10),(6.11) and (6.12) yields

$$(1+t)\|x_{S_0}\|_2^2 \leq \|x_{S_0}\|_2\sqrt{1+t}\|Ux\|_2 + \|x_{S_0}\|_2\sqrt{\delta_{2\mathbf{s}}^2 - t^2}\sum_{i=1}^{l}\sum_{j\geq 1}\|x_{S_j^i}\|_2.$$

Dividing by $\|x_{S_0}\|_2$ and employing (6.13) yields

$$(1+t)\|x_{S_0}\|_2 \leq \sqrt{1+t}\|Ux\|_2 + \sqrt{\delta_{2\mathbf{s}}^2 - t^2}\left(\frac{1}{\min\sqrt{s_i}}\|x_{S_0^c}\|_1 + \frac{1}{4}\sqrt{l}\|x_{S_0}\|_2\right). \tag{6.14}$$

Let $g(t) := \frac{\delta_{2\mathbf{s}}^2 - t^2}{(1+t)^2}$ for $|t| \leq \delta_{2\mathbf{s},\mathbf{M}}$. It is clear that $g(\delta_{2\mathbf{s},\mathbf{M}}) = g(-\delta_{2\mathbf{s},\mathbf{M}}) = 0$. Furthermore, $g(t) \geq 0$ and g is differentiable. Therefore g attains its maximum at $t_{\max}$, where $g'(t_{\max}) = 0$. A simple calculation shows us that $t_{\max} = -\delta_{2\mathbf{s},\mathbf{M}}^2$. Thus

$$g(t) \leq \frac{\delta_{2\mathbf{s},\mathbf{M}}^2 - \delta_{2\mathbf{s},\mathbf{M}}^4}{\left(1-\delta_{2\mathbf{s},\mathbf{M}}^2\right)^2} = \frac{\delta_{2\mathbf{s},\mathbf{M}}^2}{1-\delta_{2\mathbf{s},\mathbf{M}}^2}.$$

Additionally,

$$\frac{1}{\sqrt{1+t}} \leq \frac{1}{\sqrt{1-\delta_{2\mathbf{s},\mathbf{M}}}}.$$

Combining this with (6.14) yields

$$\begin{aligned}\|x_{S_0}\|_2 &\leq \frac{1}{\sqrt{1+t}}\|Ux\|_2 + \sqrt{g(t)}\left(\frac{1}{\min\sqrt{s_i}}\|x_{S_0^c}\|_1 + \frac{1}{4}\sqrt{l}\|x_{S_0}\|_2\right) \\ &\leq \frac{1}{\sqrt{1-\delta_{2\mathbf{s},\mathbf{M}}}}\|Ux\|_2 + \frac{\delta_{2\mathbf{s},\mathbf{M}}}{\sqrt{1-\delta_{2\mathbf{s},\mathbf{M}}^2}}\left(\frac{1}{\min\sqrt{s_i}}\|x_{S_0^c}\|_1 + \frac{1}{4}\sqrt{l}\|x_{S_0}\|_2\right)\end{aligned}$$

A simple rearrangement gives

$$\|x_{S_0}\|_2 \le \frac{\sqrt{1+\delta_{\mathbf{2s},\mathbf{M}}}}{\sqrt{1-\delta_{\mathbf{2s},\mathbf{M}}^2}-\delta_{\mathbf{2s},\mathbf{M}}\sqrt{l}/4}\|Ux\|_2 + \frac{\delta_{\mathbf{2s},\mathbf{M}}}{\min\sqrt{s_i}\left(\sqrt{1-\delta_{\mathbf{2s},\mathbf{M}}^2}-\delta_{\mathbf{2s},\mathbf{M}}\sqrt{l}/4\right)}\|x_{S_0^c}\|_1 \tag{6.15}$$

provided

$$\sqrt{1-\delta_{\mathbf{2s},\mathbf{M}}^2}-\delta_{\mathbf{2s},\mathbf{M}}\sqrt{l}/4 > 0. \tag{6.16}$$

Multiplying (6.15) by $\sqrt{\tilde{s}}$ yields

$$\begin{aligned}\sqrt{\tilde{s}}\|x_{S_0}\| &\le \sqrt{\tilde{s}}\frac{\sqrt{1+\delta_{\mathbf{2s},\mathbf{M}}}}{\sqrt{1-\delta_{\mathbf{2s},\mathbf{M}}^2}-\delta_{\mathbf{2s},\mathbf{M}}\sqrt{l}/4}\|Ux\|_2 + \frac{\delta_{\mathbf{2s},\mathbf{M}}\sqrt{\tilde{s}}}{\min\sqrt{s_i}\left(\sqrt{1-\delta_{\mathbf{2s},\mathbf{M}}^2}-\delta_{\mathbf{2s},\mathbf{M}}\sqrt{l}/4\right)}\|x_{S_0^c}\|_1 \\
&\le \tau\sqrt{\tilde{s}}\,\|Ux\|_2 + \frac{\delta_{\mathbf{2s},\mathbf{M}}}{\sqrt{1-\delta_{\mathbf{2s},\mathbf{M}}^2}-\delta_{\mathbf{2s},\mathbf{M}}\sqrt{l}/4}\sqrt{\sum_{k=1}^{l}\frac{s_k}{\min s_i}}\|x_{S_0^c}\|_1 \\
&\le \tau\sqrt{\tilde{s}}\,\|Ux\|_2 + \frac{\delta_{\mathbf{2s},\mathbf{M}}\sqrt{l\eta_{\mathbf{s},\mathbf{M}}}}{\sqrt{1-\delta_{\mathbf{2s},\mathbf{M}}^2}-\delta_{\mathbf{2s},\mathbf{M}}\sqrt{l}/4}\|x_{S_0^c}\|_1\end{aligned}$$

where

$$\tau = \frac{\sqrt{1+\delta_{\mathbf{2s},\mathbf{M}}}}{\sqrt{1-\delta_{\mathbf{2s},\mathbf{M}}^2}-\delta_{\mathbf{2s},\mathbf{M}}\sqrt{l}/4}.$$

It is clear that (6.8) is satisfied if condition (6.16) holds and

$$\frac{\delta_{\mathbf{2s},\mathbf{M}}\sqrt{l\eta_{\mathbf{s},\mathbf{M}}}}{\sqrt{1-\delta_{\mathbf{2s},\mathbf{M}}^2}-\delta_{\mathbf{2s},\mathbf{M}}\sqrt{l}/4} < 1. \tag{6.17}$$

A simple rearrangement shows that (6.17) is equivalent to

$$\delta_{\mathbf{2s},\mathbf{M}} < \frac{1}{\sqrt{l\left(\sqrt{\eta_{\mathbf{s},\mathbf{M}}}+\frac{1}{4}\right)^2+1}} \tag{6.18}$$

whilst (6.16) is equivalent to $\delta_{\mathbf{2s},\mathbf{M}} < \frac{1}{\sqrt{\frac{l}{16}+1}}$. Since

$$\frac{1}{\sqrt{l\left(\sqrt{\eta_{\mathbf{s},\mathbf{M}}}+\frac{1}{4}\right)^2+1}} \le \frac{1}{\sqrt{\frac{l}{16}+1}}.$$

it will suffice for (6.18) to hold, completing the proof. $\square$

6.3 Proof of Theorem 4.5 and 4.6

Proof of Theorem 4.5. The ideas behind the counterexample in this proof are similar to those in [6]. We prove this theorem in three stages. First we shall construct the matrix U. Next we shall show that our construction does indeed have a RIP in levels constant satisfying (4.5). Finally, we shall explain why z^1 exists.

Step I: Set $n = C + C^2$, where the non-negative integer C is much greater than a (we shall give a precise choice of C later). Let $x^1 \in \mathbb{C}^n$ be the vector

$$x^1 := \lambda(\underbrace{C, C, \ldots, C}_{C}, \underbrace{1, 1, \ldots, 1}_{C^2}).$$

With this definition, the first C elements of x^1 have value $C\lambda$ and C components of value 0 and the next C^2 elements have value λ. Our $(\mathbf{s},\mathbf{M})$ sparsity pattern is given by $\mathbf{s} = (1, C^2)$ and $\mathbf{M} = (0, C, C+C^2)$. Clearly, by the definition of the ratio constant, $\eta_{\mathbf{s},\mathbf{M}} = C^2$ (in particular, $\eta_{\mathbf{s},\mathbf{M}}$ is finite). Furthermore, λ is given by

$$\lambda = \frac{1}{\sqrt{C^3+C^2}} \tag{6.19}$$

so that $\|x^1\|_2 = 1$. By using this fact, we can form an orthonormal basis of $\mathbb{C}^{C+C^2}$ that includes x^1. We can write this basis as $(x^i)_{i=1}^{C+C^2}$. Finally, for a vector $v \in \mathbb{C}^{C+C^2}$, we define the linear map U by

$$Uv := \sum_{i=2}^{C+C^2} v^i x^i \text{ where } v = \sum_{i=1}^{C+C^2} v^i x^i$$

In particular, notice that the nullspace of U is precisely the space spanned by x^1, and that $v^i = \langle v, x^i \rangle$.

Step II: Let γ be an $(a\mathbf{s}, \mathbf{M})$ sparse vector. Our aim will be to estimate $\big| \|U\gamma\|_2^2 - \|\gamma\|_2^2 \big|$. Clearly, $\|U\gamma\|_2^2 - \|\gamma\|_2^2 = -|\gamma^1|^2$, where γ^1 is the coefficient corresponding to x^1 in the expansion of γ in the basis (x^i). We therefore will only need to bound $|\gamma^1| = |\langle \gamma, x^1 \rangle|$. Let S be the support of γ. Then

$$|\langle \gamma_S, x^1 \rangle| = |\langle \gamma, x_S^1 \rangle| \le \|\gamma\|_2 \|x_S^1\|_2 \le \lambda \|\gamma\|_2 \sqrt{aC^2 + C^2}$$

where we have used the Cauchy-Schwarz inequality in the second line and in the third line we have used the fact that x_S^1 has at most a elements of size λC and at most C^2 elements of size λ. From (6.19) we get

$$|\langle \gamma, x^1 \rangle| \le \sqrt{\frac{a+1}{C+1}} \|\gamma\|_2.$$

Therefore,

$$\big| \|U\gamma\|_2^2 - \|\gamma\|_2^2 \big| = |\langle \gamma, x^1 \rangle|^2 \le \frac{a+1}{C+1} \|\gamma\|_2^2.$$

By the assumption that $f(x) = o(x^{\frac{1}{2}})$, we can find a $C \in \mathbb{N}$ sufficiently large so that $\frac{a+1}{C+1} \le \frac{1}{|f(C^2)|}$. Then $\delta_{a\mathbf{s}} \le \frac{1}{|f(\eta_{\mathbf{s},\mathbf{M}})|}$ as claimed.

Step III: Let

$$z^1 := (C, \underbrace{0, 0, \ldots, 0, 0}_{C-1}, \underbrace{1, 1, \ldots, 1}_{C^2}), \qquad z^2 := (0, \underbrace{C, C, \ldots, C, C}_{C-1}, \underbrace{0, 0, \ldots, 0}_{C^2}).$$

It is clear that z^1 is $(\mathbf{s}, \mathbf{M})$-sparse. Additionally, $\|z^1\|_1 = C^2 + C$ and $\|-z^2\|_1 = (C-1)C = C^2 - C$. Because

$$U(z^1 + z^2) = \frac{U(x^1)}{\lambda} = 0,$$

we have $U(-z^2) = U(z^1)$. Since the kernel of U is of dimension 1, the only vectors z which satisfy $U(z) = U(z^1)$ are $z = z^1$ and $z = -z^2$. Moreover, $\|z^1\|_1 > \|-z^2\|_1$. Consequently

$$z^1 \notin \arg\min \|z\|_1 \text{ such that } Uz = Uz^1$$

so that recovering $(\mathbf{s}, \mathbf{M})$-sparse vectors is not guaranteed. □

Proof of Theorem 4.6. The proof of this theorem is almost identical to that of Theorem 4.5, so we shall omit details here. Again, we set x^1 so that

$$x^1 := \lambda(\underbrace{C, C, \ldots, C}_{C}, \underbrace{1, 1, \ldots, 1}_{C^2})$$

where $C \gg a$. We choose λ so that $\|x^1\|_2 = 1$. In contrast to the proof of Theorem 4.5, we take

$$\mathbf{s} = (\underbrace{1, 1, 1, \ldots, 1}_{C^2+1}), \qquad \mathbf{M} = (0, C, C+1, \ldots, C+C^2-1, C+C^2).$$

This time, there are $C^2 + 1$ levels and the ratio constant $\eta_{\mathbf{s},\mathbf{M}}$ is equal to 1. Once again, we produce an orthonormal basis of $\mathbb{C}^{C+C^2}$ that includes x^1, which we label $(x^i)_{i=1}^{C+C^2}$ and we define the linear map U by

$$Uv := \sum_{i=2}^{C+C^2} v^i x^i \text{ where } v = \sum_{i=1}^{C+C^2} v^i x^i.$$

The same argument as before proves that for any $(a\mathbf{s}, \mathbf{M})$-sparse γ,

$$\big|\,\|U\gamma\|_2^2 - \|\gamma\|_2^2\,\big| \leq \frac{a+1}{C+1}\|\gamma\|_2^2$$

and again, taking C sufficiently large so that $\frac{a+1}{C+1} \leq \frac{1}{|f(C^2+1)|}$ yields $\delta_{a\mathbf{s},\mathbf{M}} \leq \frac{1}{|f(l)|}$. The proof of the existence of z^1 is the identical to Step III in the proof of Theorem 4.5. $\square$

6.4 Proof of Theorem 4.7

Proof. Once again, we prove this theorem in three stages. First we shall construct the matrix U_2. Next, we shall show that the matrix U_2 has a sufficiently small RIP_L constant. Finally, we shall explain why both z^1 and z exist.

Step I: Let x^1 be the vector

$$x^1 := \lambda(\underbrace{0, 0, \ldots, 0}_{C^2}, \underbrace{1, 1, \ldots, 1}_{\omega(\rho,C)+1})$$

where $\omega(\rho, C) = \mathrm{ceil}(\frac{2C}{\rho})$ with a $\rho \in (0, 1)$ which we will specify later and $\mathrm{ceil}(a)$ denotes the smallest integer greater than or equal to a. In other words, the first C^2 elements of x^1 have value 0 and the next $\omega(\rho, C) + 1$ elements have value λ. We choose λ so that $\|x^1\|_2 = 1$ and C so that $C^2 > \omega(\rho, C)$. Our $(\mathbf{s}, \mathbf{M})$ sparsity pattern is given by $\mathbf{s} = (C^2, 1)$ and $\mathbf{M} = (0, C^2, C^2 + \omega(\rho, C) + 1)$. By the definition of the ratio constant, $\eta_{\mathbf{s},\mathbf{M}} = C^2$ (in particular, $\eta_{\mathbf{s},\mathbf{M}}$ is finite). Because $\|x^1\|_2 = 1$, we can form an orthonormal basis of $\mathbb{C}^{C^2+\omega(\rho,C)+1}$ that includes x^1. We can write this basis as $(x^i)_{i=1}^{C^2+\omega(\rho,C)+1}$. Finally, for a vector $v \in \mathbb{C}^{C^2+\omega(\rho,C)+1}$, we define the linear map U_2 by

$$U_2 v := \frac{\sqrt{2}w}{\tau} \text{ where } w = \sum_{i=2}^{C^2+\omega(\rho,C)+1} v^i x^i \text{ and } v = v^1 x^1 + w$$

In particular, notice that the nullspace of U_2 is precisely the space spanned by x^1, and that $v^i = \langle v, x^i \rangle$.

Step II: Let γ be an $(a\mathbf{s}, \mathbf{M})$ sparse vector. For our purposes, we take $\tau = \sqrt{2}$. Then

$$\|U_2\gamma\|_2^2 - \|\gamma\|_2^2 = -|\gamma^1|^2,$$

where γ^1 is the coefficient corresponding to x^1 in the expansion of γ in the basis (x^i). As in the proof of Theorem 4.5, $|\gamma^1| = |\langle \gamma, x^1 \rangle|$. Let S be the support of γ. Then

$$|\langle \gamma_S, x^1 \rangle| = |\langle \gamma, x^1_S \rangle| \leq \|\gamma\|_2 \|x^1_S\|_2 \leq \lambda \|\gamma\|_2 \sqrt{a}$$

where we have used the Cauchy-Schwarz inequality in the second line and in the third line we have used the fact that x^1_S has at most a elements of size λ. It is easy to see that $\lambda = \frac{1}{\sqrt{\omega(\rho,C)+1}}$. Therefore,

$$\big|\,\|U_2\gamma\|_2^2 - \|\gamma\|_2^2\,\big| = |\langle \gamma, x^1 \rangle|^2 \leq \frac{a}{\omega(\rho, C) + 1}\|\gamma\|_2^2 \leq \frac{\rho a}{2C}.$$

because $\omega(\rho, C) \geq \frac{2C}{\rho}$. By the assumption that $g(\eta_{\mathbf{s},\mathbf{M}}) \leq \frac{1}{A}\sqrt{\eta_{\mathbf{s},\mathbf{M}}}$ for some $A > 0$ and $\eta_{\mathbf{s},\mathbf{M}}$ sufficiently large, and the fact that $\eta_{\mathbf{s},\mathbf{M}} = C^2$, we must have $\frac{A}{C} \leq \frac{1}{g(\eta_{\mathbf{s},\mathbf{M}})}$. If we take ρ sufficiently small and C sufficiently large, then

$$\delta_{a\mathbf{s},\mathbf{M}} < \frac{\rho a}{2C} \leq \frac{A}{C} \leq \frac{1}{g(\eta_{\mathbf{s},\mathbf{M}})}.$$

as claimed.

Step III: Let $z^1 := x^1$ and $z := 0$, where z is defined so that $z \in \mathbb{C}^{C^2+\omega(\rho,C)+1}$. Because x^1 is in the kernel of U, $U(z - z^1) = 0$. Furthermore, it is obvious that $\|z\|_1 \leq \|z^1\|_1$. Additionally, $\|z^1\|_2 = 1$ and

$$\frac{\sigma_{\mathbf{s},\mathbf{M}}(z^1)_1}{\sqrt{\tilde{s}}} = \lambda\frac{\omega(\rho, C)}{\sqrt{C^2 + 1}} \leq \frac{\omega(\rho, C)}{\sqrt{\omega(\rho, C)\,(C^2 + 1)}} \leq \sqrt{\frac{2C + 1}{\rho\,(C^2 + 1)}} \leq \sqrt{\frac{3}{\rho\sqrt{\eta_{\mathbf{s},\mathbf{M}}}}}$$

since $\eta_{\mathbf{s},\mathbf{M}} = C^2$. Because $f(\eta_{\mathbf{s},\mathbf{M}}) = o(\eta_{\mathbf{s},\mathbf{M}}^{\frac{1}{4}})$,

$$\frac{\sigma_{\mathbf{s},\mathbf{M}}(z^1)_1}{\sqrt{\tilde{s}}} f(\eta_{\mathbf{s},\mathbf{M}}) \to 0, \quad \eta_{\mathbf{s},\mathbf{M}} \to \infty.$$

The desired result follows by taking $\eta_{\mathbf{s},\mathbf{M}}$ sufficiently large so that

$$\frac{\sigma_{\mathbf{s},\mathbf{M}}(z^1)_1}{\sqrt{\tilde{s}}} f(\eta_{\mathbf{s},\mathbf{M}}) \leq 1.$$

□

Proof of part 2. The proof of part 2 follows with a few alterations to the previous case. We now use the sparsity pattern

$$\mathbf{s} = (\underbrace{1, 1, 1, \dots, 1}_{C^2}, 1) \quad \text{and} \quad \mathbf{M} = (0, 1, 2, \dots, C^2, C^2 + \omega(\rho, C) + 1).$$

In this case, $\eta_{\mathbf{s},\mathbf{M}} = 1$ and $l = C^2 + 1$. The result follows by simply employing the same matrix U_2 with this new sparsity pattern. □

6.5 Proofs of Theorem 4.12

The counterexample for Theorem 4.12 is the same as the one used in the proof of Theorem 4.7. In that case, the matrix depended on three parameters: C, τ and ρ. We show that U_2 satisfies the ℓ^2 robust nullspace property of order $(\mathbf{s}, \mathbf{M})$ with parameters ρ and τ. The existence of z^1 and z is identical to Step III in the proof of Theorem 4.7.

Proof of part 1. Firstly, if $S \subset T$ then $\|v_S\|_2 \leq \|v_T\|_2$ and

$$\frac{\rho}{\sqrt{\tilde{s}}}\|v_{S^c}\|_1 + \tau\|Uv\|_2 \geq \frac{\rho}{\sqrt{\tilde{s}}}\|v_{T^c}\|_1 + \tau\|Uv\|_2$$

so it will suffice to prove that U_2 satisfies (4.3) for $(\mathbf{s}, \mathbf{M})$-sparse sets S with $|S| = \tilde{s}$.
Because w_S and w_{S^c} have disjoint support, by the Cauchy-Schwarz inequality applied to the vectors $(1, 1)$ and $(\|w_S\|_2, \|w_{S^c}\|_2)$ we get $\sqrt{2}\|w\|_2 \geq \|w_S\|_2 + \|w_{S^c}\|_2$. Therefore,

$$\tau\|Uv\|_2 \geq \sqrt{2}\|w\|_2 \geq \|w_S\|_2 + \|w_{S^c}\|_2. \tag{6.20}$$

Furthermore, because $|S| = \tilde{s} \geq |S^c|$ and $\rho \in (0, 1)$

$$\|w_{S^c}\|_2 \geq \frac{1}{\sqrt{|S^c|}}\|w_{S^c}\|_1 \geq \frac{1}{\sqrt{\tilde{s}}}\|w_{S^c}\|_1 \geq \frac{\rho}{\sqrt{\tilde{s}}}\|w_{S^c}\|_1 \tag{6.21}$$

Combining 6.20 and (6.21) gives

$$\begin{aligned}\tau\|Uv\|_2 + \frac{\rho}{\sqrt{\tilde{s}}}\|v_{S^c}\|_1 &\geq \|w_S\|_2 + \frac{\rho}{\sqrt{\tilde{s}}}\|w_{S^c}\|_1 + \frac{\rho}{\sqrt{\tilde{s}}}\|v_{S^c}\|_1 \\ &\geq \|w_S\|_2 + \frac{\rho}{\sqrt{\tilde{s}}}\|v_{S^c} - w_{S^c}\|_1 \\ &\geq \|w_S\|_2 + \frac{\rho}{\sqrt{\tilde{s}}}\|v^1 x^1_{S^c}\|_1 \end{aligned} \tag{6.22}$$

We shall now look at bounding $\|v^1 x^1_S\|_2$ in terms of $\|v^1 x^1_{S^c}\|_1$. We have

$$\|v^1 x^1_S\|_2 \leq \lambda|v^1| \tag{6.23}$$

since at most one element of x^1_S is non-zero and its value will be at most λ. Additionally, since each element of $x^1_{S^c}$ has value λ and there are at least $\frac{2C}{\rho}$ of them

$$\rho\|v^1 x^1_{S^c}\|_1 \geq \rho|v^1|\|x^1_{S^c}\|_1 \geq \frac{2\lambda C}{\rho}\rho|v^1| \geq 2\lambda C|v^1|.$$

Therefore,

$$\frac{\rho}{\sqrt{\tilde{s}}}\|v^1 x^1_{S^c}\|_1 \geq \frac{2\lambda C}{\sqrt{C^2+1}}|v^1| \geq \lambda|v^1|. \tag{6.24}$$

Using (6.23) and (6.24), we have $\|v^1 x^1_S\|_2 \leq \frac{\rho}{\sqrt{\tilde{s}}}\|v^1 x^1_{S^c}\|_1$. We can conclude the proof that U satisfies the ℓ^2 robust nullspace property by combining this result with (6.22) as follows:

$$\|v_S\|_2 \leq \|v^1 x^1_S\|_2 + \|w_S\|_2 \leq \frac{\rho}{\sqrt{\tilde{s}}}\|v^1 x^1_{S^c}\|_1 + \|w_S\|_2 \leq \tau\|Uv\|_2 + \frac{\rho}{\sqrt{\tilde{s}}}\|v_{S^c}\|_1.$$

□

Proof of part 2. The proof of part 2 is identical. We simply adapt the sparsity pattern so that

$$\mathbf{s} = (\underbrace{1, 1, 1, \ldots, 1}_{C^2}, 1)$$

and

$$\mathbf{M} = (0, 1, 2, \ldots, C^2, C^2 + \omega(\rho, C) + 1).$$

We can apply the proceeding argument with this new sparsity pattern to obtain the required result. □

References

[1] B. Adcock, A. C. Hansen, C. Poon, and B. Roman. Breaking the coherence barrier: A new theory for compressed sensing. *CoRR*, abs/1302.0561, 2013.

[2] B. Adcock, A. C. Hansen, B. Roman, and G. Teschke. Generalized sampling: stable reconstructions, inverse problems and compressed sensing over the continuum. *Adv. in Imag. and Electr. Phys. (to appear)*, 2013.

[3] J. Andersson and J.-O. Strömberg. On the theorem of uniform recovery of structured random matrices. *CoRR*, abs/1206.5986, 2012.

[4] T. T. Cai, L. Wang, and G. Xu. New bounds for restricted isometry constants. *IEEE Trans. Inform. Theory*, 56(9):4388–4394, 2010.

[5] T. T. Cai, L. Wang, and G. Xu. Shifting inequality and recovery of sparse signals. *IEEE Trans. Signal Process.*, 58(3, part 1):1300–1308, 2010.

[6] T. T. Cai and A. Zhang. Sharp RIP bound for sparse signal and low-rank matrix recovery. *Appl. Comput. Harmon. Anal.*, 35(1):74–93, 2013.

[7] E. J. Candès. An introduction to compressive sensing. *IEEE Signal Process. Mag.*, 25(2):21–30, 2008.

[8] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *C. R. Math. Acad. Sci. Paris*, 346(9-10):589–592, 2008.

[9] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[10] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.

[11] E. J. Candès and T. Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.

[12] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok. Introduction to compressed sensing. In *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2011.

[13] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[14] D. L. Donoho and J. Tanner. Counting faces of randomly-projected polytopes when the projection radically lowers dimension. *J. Amer. Math. Soc.*, 22(1):1–53, 2009.

[15] Y. C. Eldar and G. Kutyniok, editors. *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[16] M. Fornasier and H. Rauhut. Compressive sensing. In *Handbook of Mathematical Methods in Imaging*, pages 187–228. Springer, 2011.

[17] S. Foucart. A note on guaranteed sparse recovery via ℓ_1-minimization. *Appl. Comput. Harmon. Anal.*, 29(1):97–103, 2010.

[18] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, 2013.

[19] A. K. Louis, P. Maaß, and A. Rieder. *Wavelets*. Teubner Studienbücher Mathematik. [Teubner Mathematical Textbooks]. B. G. Teubner, Stuttgart, second edition, 1998. Theorie und Anwendungen. [Theory and applications].

[20] B. Roman, B. Adcock, and A. C. Hansen. On asymptotic structure in compressed sensing. *CoRR*, abs/1406.4178, 2014.

[21] A. M. Tillmann and M. E. Pfetsch. The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing. *IEEE Trans. Inform. Theory*, 60(2):1248–1259, 2014.